AI-ASSISTED GAMIFIED TOOL FOR BLOOM'S TAXONOMY-BASED MATHEMATICS ASSESSMENT DESIGN

Siti Sabariah Abas, Nazirah Abd Hamid, Wan Malini Wan Isa, Mumtazimah Mohamad, Maizan Mat Amin

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 22200 Besut, Terengganu, Malaysia

ABSTRACT

Artificial intelligence (AI) and gamification are increasingly integrated into multimedia educational applications to support assessment design in mathematics, particularly when guided by cognitive frameworks such as Bloom's Taxonomy. However, many existing tools emphasise static question generation or learner engagement, with limited focus on structured assessment design and user-perceived application effectiveness. This study aims to explore the perceived effectiveness of an AI-assisted gamified mathematics tool developed to support taxonomy-aligned assessment design. An exploratory research design was employed using a focus group approach involving five expert participants from higher education institutions with mathematics-related teaching experience. Data were collected through a structured questionnaire comprising demographic items and ten five-point Likert-scale statements, followed by a short focus group discussion. Descriptive statistical analysis using mean values was applied. The results indicate consistently positive perceptions of application effectiveness, with mean scores ranging from 4.0 to 4.6, particularly for efficiency, clarity of assessment design, and support for Bloom's Taxonomy–aligned question development.

KEYWORDS

Artificial Intelligent, Gamification, Bloom's Taxonomy, Mathematics Assessment, Focus Group Evaluation

1. INTRODUCTION

The integration of artificial intelligence (AI) into digital assessment systems has transformed how assessments are designed, delivered, and experienced in multimedia applications. Traditional digital assessments typically rely on static question banks and fixed scoring mechanisms, which limit their ability to adapt to learners' behaviours or provide meaningful feedback. In contrast, AI-assisted assessment tools can support adaptive question selection, real-time feedback, and intelligent scaffolding based on user interaction patterns. Prior research highlights that AI, when used as an assistive component rather than a replacement for human judgement, can enhance the effectiveness and flexibility of assessment systems in technology-enhanced learning environments [1].

Gamification has also played a significant role in reshaping assessment design within digital and multimedia applications. By incorporating game elements such as points, progress indicators, challenges, and immediate feedback, gamified assessments aim to reduce test anxiety and increase user engagement. Empirical studies have shown that gamification can improve learners' motivation and willingness to participate in assessment activities, particularly in digital

environments where disengagement is a common challenge [2], [3]. In assessment contexts, gamification shifts the focus from purely evaluative tasks to more interactive and experience orientated activities, which can encourage sustained user interaction.

The combination of AI and gamification offers new opportunities for assessment design in multimedia systems. AI can dynamically support gamified assessment by adjusting difficulty levels, recommending suitable tasks, or providing personalised hints; while gamification enhances user experience through interactive and visually engaging elements. Research suggests that such hybrid approaches can lead to more meaningful assessment experiences by balancing engagement with cognitive challenge [4]. However, studies also caution that, without clear pedagogical alignment, AI-driven and gamified assessments may prioritise engagement over learning quality.

In mathematics education, assessment design presents additional challenges due to the need to evaluate both procedural accuracy and conceptual understanding. Gamified assessment tools can make mathematical problem-solving more approachable, but they risk reinforcing surface-level performance if not carefully structured. AI-assisted assessment design provides a mechanism to align tasks with varying cognitive demands, enabling the assessment of different levels of understanding. As highlighted in prior work, effective assessment design must be grounded in learning frameworks that guide task complexity and cognitive expectations [5], [6]. This study situates AI-assisted gamified assessment within such a framework to explore its potential effectiveness in a multimedia mathematics application.

Mathematics assessment presents unique challenges due to the abstract nature of mathematical concepts and the need to evaluate both procedural skills and conceptual understanding. While digital applications offer advantages such as automation and immediate feedback, many existing mathematics apps primarily focus on rote practice and surface-level performance indicators [7]. This limitation reduces their ability to assess deeper cognitive processes such as application and problem-solving. Another challenge lies in the alignment between assessment tasks and learning objectives. Research suggests that digital mathematics assessments often lack clear pedagogical grounding, resulting in activities that measure speed or correctness without capturing reasoning processes [8]. In gamified environments, this issue can be amplified when scoring and rewards overshadow learning goals. Consequently, the effectiveness of digital assessments depends not only on technology but also on how assessment tasks are designed.

Additionally, user experience plays a critical role in digital mathematics assessment. Poorly designed interfaces or overly complex interactions can distract learners from the assessment task itself. Multimedia applications must balance visual appeal with cognitive clarity to avoid increasing cognitive load unnecessarily [6]. These challenges highlight the need for structured frameworks that guide assessment design while leveraging the strengths of multimedia systems.

Bloom's Taxonomy has long been used as a framework for categorising educational objectives and assessing cognitive processes. The revised taxonomy by [5] organises cognitive skills from lower-order processes such as remembering and understanding to higher-order processes such as analysing and evaluating. This structure provides a systematic way to design assessment tasks that target different levels of cognitive level. In digital and multimedia learning contexts, Bloom's Taxonomy has been applied to evaluate whether learning activities promote meaningful cognitive engagement. Studies have shown that many digital tools predominantly target lower-order cognitive skills, particularly in automated assessment environments [9]. Applying Bloom's Taxonomy in assessment design helps ensure that tasks move beyond memorisation and encourage deeper thinking.

For gamified and AI-assisted applications, Bloom's Taxonomy offers a practical framework to align game mechanics and assessment tasks with cognitive objectives. By mapping assessment items to specific taxonomy levels, designers can create balanced assessments that support progressive learning. Previous research suggests that systems explicitly aligned with Bloom's Taxonomy are more likely to support effective learning experiences [10]. This study adopts Bloom's Taxonomy as a guiding framework for assessment design within a gamified multimedia application.

The motivation for integrating AI into assessment tools stems from the limitations of static digital assessments. Traditional digital assessments often rely on predefined questions and fixed feedback, which may not address individual learner needs. AI-assisted assessment tools offer the potential to support users by providing adaptive feedback, content suggestions, or scaffolding during task completion [1]. In multimedia learning environments, AI can enhance assessments by analysing user interactions and guiding learners through assessment tasks. Rather than replacing human judgement, AI in this context functions as an assistive component that supports assessment design and delivery. Research indicates that AI-assisted systems can improve user experience and engagement when used to augment, rather than automate, assessment processes [11].

From a design perspective, AI-assisted tools also enable more flexible assessment structures. For example, AI can help categorise questions according to difficulty levels or cognitive objectives, supporting alignment with frameworks such as Bloom's Taxonomy. This capability is particularly relevant in exploratory studies, where the goal is to evaluate how users perceive and interact with AI-supported assessment features within multimedia applications.

The primary objective of this study is to explore the perceived effectiveness of an AI-assisted, gamified mathematics tool designed for assessment based on Bloom's Taxonomy. Rather than seeking statistical validation, the study adopts a qualitative focus group approach to gather in depth feedback on user experience, engagement, and perceived learning support. This approach is appropriate for early-stage evaluation of multimedia applications. The scope of the study is limited to exploratory evaluation using a small focus group (N = 5). Such an approach allows for detailed discussion and observation of user interactions with the application. Previous multimedia research has demonstrated that focus group studies are suitable for identifying usability issues, design strengths, and areas for improvement in interactive systems [12], [13].

Contributions of this study are threefold. First, unlike prior AI-assisted or gamified mathematics applications that primarily emphasise learner engagement or automated assessment, this study focuses explicitly on assessment design, guided by Bloom's Taxonomy. Second, the study demonstrates how Bloom's Taxonomy can be operationalised within an AI-assisted gamified tool to support structured question construction across cognitive levels, addressing a gap identified in prior digital assessment research. Third, this work contributes user-centred empirical evidence through expert-based focus group evaluation, complementing existing outcome-focused or system-centric studies by examining perceived effectiveness from an assessment design perspective.

2. RELATED WORK

This section reviews prior research on AI-assisted assessment systems, gamified mathematics applications, Bloom's Taxonomy in digital assessment design, and user-centred evaluation approaches for multimedia systems. While existing studies highlight the growing use of AI and gamification in education, they reveal limited integration of cognitive frameworks such as

Bloom's Taxonomy and a lack of qualitative, user-focused evaluation of assessment design effectiveness.

2.1. AI-Assisted Tools in Educational Assessment

Artificial intelligence has increasingly been integrated into educational assessment systems to enhance efficiency, adaptability, and feedback quality. Early AI-based assessment tools focused on automated scoring and rule-based feedback, particularly for structured question formats. More recent systems incorporate machine learning and learning analytics to support adaptive assessments, personalised feedback, and assessment management within digital platforms [14], [15]. These developments reflect a shift from static assessment delivery toward intelligent, user responsive systems.

Recent systematic reviews confirm that intelligent assessment and management constitute one of the core application areas of AI in education. [14] analysed a large body of AI-in-education literature and reported that assessment-related applications are among the fastest-growing research themes, particularly in higher education. Similarly, [16] highlighted that AI is increasingly used to support formative assessment processes, enabling real-time feedback and adaptive task sequencing rather than replacing human evaluators.

However, researchers also caution against the uncritical adoption of AI in assessment. [17] argue that AI-assisted assessment must be understood as part of a broader socio-technical system that includes pedagogical intent, institutional governance, and learner trust. Recent higher-education reviews emphasise that AI is most effective when positioned as an assistive tool that augments assessment design and delivery, rather than an autonomous decision-maker [18]. This perspective supports the use of qualitative, user-centred evaluations such as focus groups to explore how users experience AI-assisted tools in designing assessments.

2.2. Gamified Applications for Mathematics Learning and Assessment

Gamification has been widely adopted in mathematics learning applications to address challenges related to learner motivation, engagement, and persistence. Game elements such as points, progress indicators, challenges, and immediate feedback are frequently embedded into mathematics apps to encourage repeated practice and sustained interaction. Empirical studies consistently report positive effects of gamification on engagement and learner attitudes toward mathematics, particularly in digital contexts [19], [20].

Despite these motivational benefits, recent literature highlights mixed findings regarding learning effectiveness. A systematic review of learning games in K–12 mathematics education found that while gamified systems often improve engagement, their impact on conceptual understanding and higher-order thinking varies considerably across studies [21]. This variability is often attributed to differences in instructional design quality and the degree to which mathematical content is meaningfully integrated into gameplay mechanics.

More recent mathematics-specific reviews reinforce this concern. [22] reported that many gamified mathematics applications prioritise surface-level interactions, such as speed and reward accumulation, rather than cognitive challenge. Broader education-wide reviews published between 2023 and 2024 similarly conclude that gamification is most effective when aligned with learning objectives and cognitive frameworks rather than implemented as an add-on feature [16]. These findings highlight the need for structured assessment design approaches when gamification is used in mathematics applications.

2.3. Bloom's Taxonomy in Digital and App-Based Assessment

Bloom's Taxonomy has long served as a foundational framework for categorising learning objectives and assessment tasks according to cognitive complexity. The revised taxonomy proposed by [5] provides a hierarchical structure that ranges from remembering and understanding to applying, analysing, evaluating, and creating. This structure has been widely adopted in instructional and assessment design to ensure balanced cognitive coverage.

In digital and multimedia learning environments, Bloom's Taxonomy has increasingly been operationalised as a design and evaluation framework. Studies have shown that many digital assessments disproportionately target lower-order cognitive skills due to ease of automation and scoring [6], [9]. Applying Bloom's Taxonomy allows designers to intentionally structure assessment tasks that address higher-order thinking, even within automated or app-based systems.

Recent app-based studies demonstrate practical applications of Bloom's Taxonomy in assessment design. [23] reported the development of a mobile learning application that explicitly supports higher-order thinking skills (HOTS) by aligning assessment questions with revised Bloom's levels.

2.4. Evaluation Approach

Evaluation of AI-assisted and gamified educational applications commonly emphasises user perception, usability, and perceived effectiveness, particularly during the early stages of system development. Rather than focusing on large-scale outcome measurement, many studies adopt user-centred evaluation approaches to examine how users experience system features, such as AI assistance, feedback mechanisms, and interactive assessment design. This is especially relevant for multimedia applications, where interaction quality and user interpretation play a central role in determining application effectiveness [13].

Questionnaire-based evaluation, particularly using Likert-scale instruments, is widely employed to capture users' perceived effectiveness and usability of educational technologies. Likert-scale questionnaires allow researchers to systematically measure subjective perceptions, such as usefulness, clarity of feedback, and perceived support provided by intelligent features. Prior studies highlight that questionnaire-based perception measures are suitable for exploratory evaluations of AI-assisted systems, where user acceptance and perceived value are critical indicators of system success [18], [24].

In parallel, focus group methods are frequently used to complement questionnaire data by providing deeper insight into users' experiences and interpretations. Focus groups enable participants to elaborate on their responses, clarify ratings, and discuss specific design features in relation to assessment tasks and cognitive alignment. [13] emphasise that focus groups are particularly effective for evaluating interactive and multimedia systems, as they reveal nuanced perceptions that may not be evident from quantitative data alone.

Recent educational technology studies further demonstrate that combining questionnaires and focus group discussions is effective for evaluating AI-assisted and gamified applications. [25] report that this mixed qualitative–descriptive approach supports a more comprehensive understanding of application effectiveness, particularly in systems involving adaptive behaviour and gamified interaction. For assessment-orientated applications grounded in Bloom's Taxonomy, such approaches are well suited to examining whether users perceive the system as supporting meaningful assessment design and cognitive engagement.

2.5. Research Gap Analysis

Despite the growing body of research on AI-assisted educational assessments, most existing studies emphasise system capabilities, automation, and learning analytics rather than users' perceptions of AI assistance in assessment design, particularly in mathematics contexts. Recent reviews confirm that intelligent assessment is a core application area of AI in education; however, empirical investigations largely focus on learning outcomes or large-scale implementations [14], [16]. There remains a lack of user-centred, qualitative studies examining how expert users experience and evaluate AI-supported tools during the assessment design process. This gap is significant, as prior work highlights that perceived usefulness, trust, and clarity of AI support are critical factors influencing adoption and effective use of AI-assisted assessment systems [17], [18].

Similarly, while gamification is widely adopted in mathematics applications, existing research predominantly targets engagement and motivation rather than assessment design quality and cognitive alignment. Systematic reviews indicate that many gamified mathematics tools prioritise surface-level interaction and reward mechanisms, often without explicit integration of cognitive frameworks such as Bloom's Taxonomy [21], [22]. Although Bloom's Taxonomy is frequently referenced in digital learning research, its operationalisation within AI-assisted, gamified assessment tools and evaluation from the users' perspective remain underexplored. Furthermore, few studies combine questionnaire-based measures of perceived effectiveness with focus group discussion to evaluate such applications. These gaps are illustrated in Figure 1, which highlights the imbalance in existing research between technology-centric approaches, emphasising system capabilities, automation, and engagement, and the limited attention given to user-centred evaluation, assessment design quality, and cognitive alignment with Bloom's Taxonomy in AI assisted, gamified mathematics assessment tools.

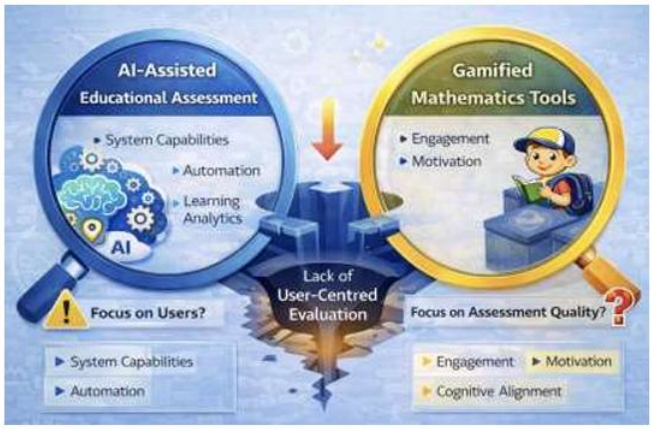

Figure 1: Research gap identified in this project

This study addresses these gaps by adopting a mixed descriptive-qualitative approach to explore the perceived application effectiveness of an AI-assisted gamified mathematics tool designed for assessment based on Bloom's Taxonomy.

3. METHOD

This chapter outlines the methodological approach adopted to evaluate the perceived effectiveness of the AI-assisted gamified mathematics tool for assessment design based on Bloom's Taxonomy. It describes the system overview, research design, participant selection, data

collection procedures, and data analysis approach employed in this exploratory focus group study.

3.1. Overview of the AI-Assisted Gamified Mathematics Tool

The AI-assisted gamified mathematics tool developed in this study is a multimedia application designed to support assessment design and interaction based on Bloom's Taxonomy. The tool integrates interactive multimedia elements (visual feedback, progress indicators, and task prompts) with gamification mechanics such as scoring, levels, and immediate feedback. These elements aim to enhance user engagement while supporting structured assessment activities rather than purely entertainment-driven interaction [3], [6].

Artificial intelligence in the tool functions as an assistive component rather than an autonomous decision-maker. Specifically, AI features were designed to support question organisation, adaptive task presentation, and feedback alignment with cognitive levels defined by Bloom's Taxonomy. This approach is consistent with recent AI-in-education literature, which emphasises augmentation of assessment processes rather than full automation [1], [14].

Gamification elements were deliberately aligned with assessment goals to avoid superficial engagement. Rather than rewarding speed alone, game mechanics were used to structure task progression and reinforce cognitive engagement. Prior research shows that gamified systems are most effective when game elements are integrated with pedagogical objectives and cognitive frameworks [22]. The tool therefore operationalises Bloom's Taxonomy as a guiding structure for assessment task design and user interaction.

For this study, the AI-assisted features of the proposed tool were designed and implemented as part of a custom multimedia application. The system does not rely on a standalone, externally deployed large language model (LLM); instead, AI assistance is implemented through rule-based logic and adaptive mechanisms that support question organisation, taxonomy-level alignment, and feedback presentation. Bloom's Taxonomy levels and associated keywords are explicitly encoded within the system to guide assessment design. Thus, the AI component functions as an assistive decision-support mechanism, augmenting educators' assessment design processes rather than autonomously generating or evaluating assessment content.

3.2. Research Design

This study adopted an exploratory research design to examine the perceived effectiveness of the AI-assisted gamified mathematics tool. Exploratory designs are appropriate when investigating early-stage applications, particularly when the objective is to gather formative feedback and refine the system design rather than produce generalisable outcomes [13].

A key research gap motivating this design lies in the limited empirical work that integrates AI assistance, gamification, and Bloom's Taxonomy specifically for assessment design within mathematics applications. Existing studies often focus on learning engagement or achievement outcomes, while assessment design processes and user perceptions remain underexplored [18], [21]. This study addresses this gap by focusing on how users experience and evaluate such a tool.

Figure 2 shows the research design framework that outlines the step-by-step process of developing and evaluating the AI-assisted gamified mathematics tool, from requirement analysis to effectiveness evaluation using expert feedback. User requirements were identified through analysis of prior literature on digital mathematics assessment and gamified learning systems. Research consistently highlights the need for clear cognitive alignment, usable interfaces, and

meaningful feedback in assessment tools [16], [23]. These requirements informed the conceptual design of the application.

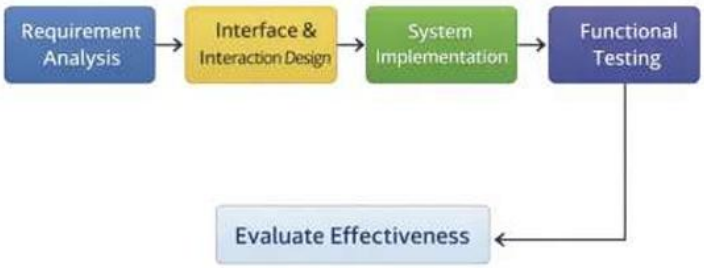

Figure 2: Research design framework

Based on the identified requirements, the application was designed, developed, and iteratively tested. The development process followed a structured flow: requirement analysis, interface and interaction design, system implementation, and functional testing. Such a staged development process is widely recommended in multimedia application research to ensure usability and functional validity before evaluation [12], [25].

3.3. Participants

Participants were recruited using purposive sampling to ensure relevant expertise for evaluating the assessment design features of the application. A focus group of five participants (N = 5) was formed, consistent with recommendations for exploratory usability and system evaluation studies [12], [13]. Small focus groups are particularly suitable for in-depth discussion and interaction based evaluation.

All participants were academic staff from a higher education institution, and the evaluation context reflects tertiary-level mathematics assessment design ensuring that feedback on assessment content and cognitive alignment was informed and meaningful. Each participant had more than five years of professional experience in mathematics teaching, assessment, or related educational roles. This criterion aligns with prior research suggesting that expert users provide higher-quality evaluative feedback for educational tools [26].

The use of expert participants supports the study's objective of assessing application effectiveness from a design and assessment perspective, rather than measuring student learning outcomes. Recent studies emphasise that expert-based evaluation is appropriate for early-stage assessment tools and contributes valuable insights into design validity and usability [18].

3.4. Data Collection Procedure

The data collection procedure was designed to systematically capture participants' perceptions and experiences when interacting with the AI-assisted gamified mathematics tool. Given the exploratory nature of the study and the focus on application effectiveness, a structured, yet user centred procedure was adopted. Such approaches are commonly recommended for evaluating early-stage multimedia and educational applications, where user interaction and immediate feedback constitute critical sources of evidence [13], [25].

The session began with a briefing session, during which participants were informed about the study objectives, the functionality of the application, and the overall evaluation process. Ethical considerations, including voluntary participation, confidentiality, and the right to withdraw, were clearly explained. This initial briefing ensured that participants shared a common understanding of the task context and evaluation goals, which is considered good practice in human–computer interaction and educational technology research [12], [27].

Following the briefing, participants proceeded to the system interaction phase, where they engaged in task-based interaction with the developed application. Each participant used the system for a predefined duration and completed guided assessment-related tasks designed to reflect realistic mathematics assessment design and interaction scenarios. This phase enabled participants to directly experience the AI-assisted features, gamification elements, and Bloom's Taxonomy-aligned assessment structure. Task-based system use is widely recognised as an effective method for eliciting authentic user responses and evaluating usability and interaction quality in multimedia applications [12], [13], [25].

Immediately after system interaction, participants completed a questionnaire employing a Likert scale format to measure their perceived effectiveness of AI assistance within the gamified mathematics assessment tool. The questionnaire focused on perceptions of guidance clarity, feedback relevance, and AI support for assessment tasks. Likert scales are commonly used in early-stage educational technology evaluations to capture subjective perceptions of assistive system effectiveness [18], [24]. Responses were collected immediately after interaction to reduce recall bias, and mean values were calculated to summarise overall trends in perceived AI effectiveness, consistent with exploratory studies involving small samples [14], [27], [28].

Following questionnaire completion, a short focus group discussion was conducted to gather clarifying feedback and elaboration on participants' responses. The discussion protocol focused on key themes such as usability, clarity of assessment tasks, perceived alignment with Bloom's Taxonomy, and the role of AI assistance. Focus group discussions complement questionnaire data by capturing nuanced perceptions that may not be fully reflected in numerical ratings, and they are widely used in exploratory multimedia evaluation studies [13].

Finally, all responses were digitally recorded and exported for analysis. Questionnaire data were stored in a structured digital format to support descriptive statistical analysis, while qualitative feedback from the focus group was retained for contextual interpretation. Systematic digital data handling is a standard practice in exploratory educational technology research to ensure data integrity, traceability, and transparency [27].

3.5. Data Analysis Approach

The data analysis employed a descriptive statistical approach to examine participants' perceptions of the effectiveness of AI assistance within the gamified mathematics assessment tool. Given the exploratory nature of the study and the small focus group sample size (N = 5), inferential statistical testing was not performed, as such methods require larger samples and assumptions that are not satisfied in small-sample evaluations. Instead, descriptive analysis was selected to summarise response patterns and provide an interpretable overview of user perceptions, which is consistent with recommended practices for early-stage educational technology and multimedia system evaluations [13], [28].

Participant responses were collected using a five-point Likert scale, where 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree. The Likert items were specifically designed to measure the perceived effectiveness of AI assistance, including the clarity of AI

generated guidance, relevance of feedback during assessment tasks, and perceived support for Bloom's Taxonomy–aligned assessment design. Likert-scale instruments are widely used to quantify subjective perceptions of system usefulness and effectiveness in educational technology research, particularly when evaluating assistive or intelligent system features [18], [24].

For each questionnaire item, mean scores were computed to represent the central tendency of participant responses. Mean values provide a concise numerical summary that facilitates comparison across evaluation dimensions and are commonly reported in exploratory studies involving small samples [28]. In this study, higher mean scores indicated stronger agreement regarding the effectiveness of AI assistance, while lower mean scores reflected weaker perceived support. The analysis focused on identifying overall trends rather than statistical significance, thereby avoiding overinterpretation of results.

The descriptive findings were interpreted alongside qualitative feedback obtained during the focus group discussion to contextualise the numerical results. This combined descriptive qualitative approach enhances interpretive validity by linking quantitative ratings with participants' explanations and observations, which is particularly important when evaluating complex multimedia systems integrating AI and gamification [13], [25]. Overall, this analysis strategy provides a technically sound and methodologically transparent basis for assessing perceived application effectiveness while remaining appropriate for the exploratory scope of the study.

The CT-based English writing module was developed based on the ADDIE model and consisted of eight thematic units. Each unit embeds CT elements such as decomposition, pattern recognition, abstraction, algorithmic thinking and logical reasoning into writing tasks as shown in Table 2. This approach enables pupils to structure their thoughts more logically and systematically.

4. RESULTS

This chapter presents the results of the exploratory evaluation of the AI-assisted gamified mathematics tool based on data collected from the focus group study. It reports the structure of the research instrument, demographic characteristics of the participants, and descriptive findings on perceived application effectiveness.

4.1. Research Instrument

The research instrument used in this study was designed to systematically capture expert users' background information and their perceptions of the effectiveness of the AI-assisted gamified mathematics tool for assessment design. The instrument consisted of two main sections: Section A (Demographics) and Section B (Application Effectiveness). This structure aligns with common practices in educational technology evaluation, where contextual user information is collected prior to effectiveness assessment [13].

Section A comprised six demographic questions aimed at profiling the participants' professional background, teaching domain, and prior exposure to gamification in assessment design. Collecting demographic data is essential in exploratory studies to contextualise responses and ensure that participants possess relevant expertise for evaluating the system [26]. In this study, Section A ensured that the results reflected expert perspectives rather than novice impressions.

Section B consisted of ten Likert-scale items designed to measure participants' perceived effectiveness of the application. The items focused on efficiency, clarity, cognitive alignment with Bloom's Taxonomy, and perceived support for assessment design. Likert-scale instruments are widely used to evaluate the perceived effectiveness of educational applications, particularly in early-stage system evaluation where subjective expert judgement is critical [18], [24].

All items in Section B were rated using a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The mean scores were calculated to summarise overall trends in perceived effectiveness, consistent with recommended practices for descriptive analysis in small sample exploratory studies [28].

4.2. Results of Demographic Analysis

The demographic analysis indicates that the participants represented a relatively homogeneous expert group, which is appropriate for a focus group–based evaluation study. Table 1 presents the homogeneous demographic profile that supports the exploratory nature of the focus group study and ensures that feedback reflects expert perspectives rather than prior familiarity with gamified assessment tools.

Table 1. Demographic Profile of Focus Group Participants (N = 5)

Variable	Category	n	%
Age Range	35–40 years	5	100%
Gender	Female	5	100%
Institutional Affiliation	Faculty of Informatics and Computing, UniSZA	5	100%
Teaching Specialisation	Mathematics	4	80%
	Statistics	1	20%
Teaching Experience	5 years	1	20%
	6 years	2	40%
	7 years	2	40%
Prior Gamification Assessment Experience	Yes	1	20%
	No	4	80%

All participants were within the age range of 35 to 40 years, suggesting a mature professional cohort with sufficient teaching and assessment experience to evaluate the application meaningfully.

In terms of gender and institutional affiliation, all participants were female academic staff from the Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA). This consistency in institutional background ensures that participants share similar academic and assessment contexts, reducing variability due to institutional policy differences. Such contextual consistency is beneficial in exploratory system evaluation studies where depth of feedback is prioritised over representativeness [13].

Regarding teaching specialisation, four participants identified mathematics as their primary teaching field, while one participant specialised in statistics. This distribution confirms that the majority of participants possessed direct relevance to mathematics assessment design while still allowing for a complementary quantitative perspective. Prior research highlights that involving closely related domains can enrich evaluation without compromising relevance [26].

Participants' teaching experience ranged from five to seven years, with one participant having five years of experience, two participants six years, and two participants seven years. This

indicates that all participants met the minimum experience threshold to be considered expert users. Interestingly, only one participant reported prior experience using gamification applications for assessment design, while four had no such experience. This finding is important, as it suggests that the perceived effectiveness results were not strongly influenced by prior familiarity with gamified assessment tools.

4.3. Results of Application Effectiveness

The results of Section B demonstrate consistently high mean scores across most application effectiveness items, indicating a positive overall perception of the AI-assisted gamified mathematics tool. The highest mean score (M = 4.6) was recorded for the item stating that the application helps users generate assessment questions more quickly. This suggests that efficiency is one of the most salient perceived benefits of the tool, aligning with prior findings that AI assisted systems can reduce cognitive and time-related workload in assessment design [14].

Most items related to cognitive alignment with Bloom's Taxonomy recorded mean scores of 4.4, including the ability to generate questions aligned with selected taxonomy levels, keyword guidance for focused question construction, and support for producing multiple question types. These results indicate strong agreement among participants that the application effectively supports structured assessment design. This is consistent with recent studies reporting that AI assisted tools can enhance clarity and consistency in taxonomy-based assessment development [23].

Items related to confidence and quality assurance, such as perceived confidence in the generated questions and support for academic assessment standards, also recorded high mean values (M = 4.4). These findings suggest that participants perceived the tool not only as efficient but also as reliable for formal academic assessment contexts. Prior literature emphasises that perceived quality and trust are critical factors influencing the adoption of AI-assisted assessment tools among educators [17], [18].

The lowest mean score was recorded for the overall effectiveness item (M = 4.0), which still indicates agreement but reflects a more cautious evaluation. This is expected in exploratory studies, where expert users may recognise both strengths and areas for improvement. Additionally, the item related to understanding Bloom's Taxonomy application recorded a slightly lower mean (M = 4.2), suggesting that while the tool supports taxonomy-based design, further refinement or instructional scaffolding may enhance conceptual clarity.

Overall, the results indicate that participants perceived the AI-assisted gamified mathematics tool as effective in supporting assessment design, particularly in terms of efficiency, cognitive alignment, and practical usability, as shown in Table 2. These findings provide formative evidence supporting the application's potential while remaining consistent with the exploratory scope of the study.

Table 2. Mean Scores, M for Application Effectiveness Items.

Dimension	Evaluation Item	Mean Score (*M*)	Interpretation
Efficiency	Helps users generate assessment questions more quickly	4.6	Very high agreement; efficiency is the most salient perceived benefit of the application
Cognitive Alignment (Bloom's Taxonomy)	Generates questions aligned with selected Bloom's Taxonomy levels	4.4	Strong agreement; supports structured taxonomy-based assessment design
	Provides keyword guidance for focused question construction	4.4	Strong agreement; facilitates clarity and precision in question formulation
	Supports generation of multiple question types	4.4	Strong agreement; enhances flexibility in assessment design
Confidence & Quality Assurance	Increases confidence in the generated assessment questions	4.4	Strong agreement; users perceive outputs as reliable for academic use
	Supports adherence to academic assessment standards	4.4	Strong agreement; reinforces perceived quality and trustworthiness
Conceptual Understanding	Enhances understanding of Bloom's Taxonomy application	4.2	High agreement; indicates effectiveness, with room for further instructional support
Overall Effectiveness	Overall effectiveness of the AI-assisted gamified tool	4.0	Moderate to high agreement; reflects a cautious but positive evaluation typical of exploratory studies

5. DISCUSSION

This study set out to explore the perceived effectiveness of an AI-assisted gamified mathematics tool designed to support assessment development based on Bloom's Taxonomy. The discussion focuses on how the results align with existing literature, how they address the identified research gaps, and what they contribute to current understanding of AI-assisted assessment tools in mathematics education.

5.1. Effectiveness of AI Assistance in Assessment Design

The results indicate that participants perceived the AI-assisted features as highly effective in supporting assessment design, particularly in terms of efficiency and task clarity. High mean scores for items related to faster question generation and reduced planning time suggest that AI assistance can alleviate the cognitive and time-related workload associated with assessment preparation. These findings are consistent with prior studies reporting that AI-supported assessment tools enhance efficiency without replacing educators' professional judgement [1], [14].

Participants also reported strong agreement regarding the relevance and contextual accuracy of AI-generated questions. This supports recent literature, emphasising that AI systems are most valuable when they assist in structuring and refining assessment tasks rather than autonomously generating content [18]. The perceived effectiveness observed in this study reinforces the view that AI should function as a design aid, particularly in cognitively structured domains such as mathematics assessment.

Importantly, the results suggest that positive perceptions were observed even though most participants had limited prior experience with gamified assessment tools. This implies that the AI-assisted features were accessible and understandable to expert users, supporting earlier

findings that usability and clarity are critical factors influencing acceptance of intelligent educational systems [17].

5.2. Role of Gamification in Supporting Assessment-Orientated Tasks

While gamification is often discussed in relation to motivation and engagement, the findings of this study suggest that gamified elements can also support assessment-orientated functions when appropriately designed. Participants' agreement that the application supported the generation of multiple question types and structured task progression indicates that gamification can facilitate systematic assessment design rather than merely enhancing surface engagement.

This finding address concerns raised in prior reviews that many gamified mathematics applications emphasise rewards and speed at the expense of cognitive depth [21], [22]. In contrast, the results of this study suggest that when gamification is explicitly aligned with assessment objectives and cognitive frameworks, it can contribute positively to structured question development.

The relatively high mean scores related to academic standards compliance further indicate that gamification, when integrated carefully, does not necessarily undermine assessment rigour. This supports broader educational technology literature arguing that gamification should be embedded as a functional design element rather than an add-on feature [16].

5.3. Bloom's Taxonomy as a Structuring Framework for AI-Assisted Assessments

One of the central contributions of this study lies in demonstrating how Bloom's Taxonomy can be operationalised within an AI-assisted, gamified assessment tool. Participants reported strong agreement that the application helped them generate questions aligned with selected taxonomy levels and supported focused question construction through taxonomy-based keywords. These findings extend prior work that primarily treats Bloom's Taxonomy as a conceptual reference rather than an actionable design mechanism in digital tools [6].

The slightly lower mean score related to understanding Bloom's Taxonomy suggests that while the tool supports taxonomy-based design, additional instructional scaffolding may further enhance users' conceptual clarity. This aligns with previous studies indicating that digital tools can support higher-order thinking only when users are adequately guided in applying cognitive frameworks [23].

Overall, the findings suggest that Bloom's Taxonomy remains a practical and effective framework for structuring assessment design within multimedia applications, particularly when supported by AI features that guide alignment and consistency.

5.4. Addressing the Identified Research Gaps

The results directly address the research gaps identified in the literature review. First, by focusing on expert users' perceptions, this study contributes user-centred evidence to a research area that is often dominated by system-centric or outcome-based evaluations. The use of a focus group combined with descriptive questionnaire analysis provides insight into how AI-assisted assessment tools are experienced in practice.

Second, the study responds to the lack of empirical work examining the integration of AI assistance, gamification, and Bloom's Taxonomy specifically for assessment design in

mathematics. Rather than measuring learning outcomes, the study evaluates application effectiveness from a design and usability perspective, which is appropriate for early-stage tool development.

Finally, the findings support the use of small-scale exploratory evaluations in educational technology research. As recommended in prior methodological literature, such studies play a crucial role in refining system design before larger-scale deployment and validation [13].

6. CONCLUSION AND FUTURE WORK

This study explored the perceived effectiveness of an AI-assisted gamified mathematics tool designed to support assessment development based on Bloom's Taxonomy. Using a focus group approach and a Likert-scale questionnaire, the study gathered expert users' perceptions regarding efficiency, cognitive alignment, and practical usability of the application. The findings indicate that participants generally perceived the tool as effective in assisting assessment design, particularly in reducing preparation time and supporting taxonomy-aligned question construction.

The results also suggest that integrating AI assistance and gamification within a structured cognitive framework can support assessment-oriented tasks rather than merely enhancing engagement. Participants reported positive perceptions even with limited prior experience using gamified assessment tools, indicating that the application design was accessible to expert users. These findings are consistent with the exploratory scope of the study and contribute formative evidence to the growing literature on AI-assisted assessment tools in mathematics education.

Despite the positive outcomes, this study has limitations that should be acknowledged. The small focus group size and single-institution context limit the generalisability of the findings. These limitations are inherent to early-stage application evaluation and align with the study's design objectives. Furthermore, this study evaluated perceived effectiveness rather than objective learning or performance outcomes. While objective measures such as mathematics test scores could provide stronger evidence of learning impact, the present study intentionally adopted a user centred exploratory evaluation to examine assessment design support at an early development stage. Such an approach is consistent with prior multimedia and AI-assisted system evaluations, where usability and perceived usefulness are examined before large-scale experimental validation.

Future work may extend this research by involving larger and more diverse participant groups, including educators from different institutions and subject areas. Further studies could also examine the use of the application in authentic classroom settings or investigate its impact on assessment quality and student learning outcomes over time. Such work would help validate and refine the tool before broader implementation. Additionally, future studies will incorporate objective performance-based measures, such as pre–post mathematics assessment scores or expert-rated assessment quality metrics, to triangulate perceived effectiveness with measurable learning and assessment outcomes.

ACKNOWLEDGEMENTS

The first author wishes to thank the Ministry of Higher Education (MOHE), Malaysia, and Universiti Sultan Zainal Abidin, Malaysia, for financial support via Internal Grant (Scholarship of Teaching and Learning - UNISZA/2024/SOTL/12 (RK086)).

REFERENCES

[1] W. Holmes, M. Bialik, and C. Fadel, Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Boston, MA, USA: Center for Curriculum Redesign, 2019.

[2] Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining gamification," in Proc. 15th Int. Academic MindTrek Conf., Tampere, Finland, 2011, pp. 9–15.

[3] W. Oliveira, J. Hamari, L. Shi, A. M. Toda, L. Rodrigues, P. T. Palomino, and S. Isotani, "Tailored gamification in education: A literature review and future agenda," Education and Information Technologies, vol. 28, no. 1, pp. 373–406, 2023.

[4] K. M. Kapp, The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education. San Francisco, CA, USA: Pfeiffer, 2012.

[5] L. W. Anderson and D. R. Krathwohl, A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York, NY, USA: Longman, 2001.

[6] R. E. Mayer, Multimedia learning, 2nd ed. Cambridge University Press, 2009.

[7] M. M. Riconscente, "Results from a controlled study of the iPad fractions game Motion Math," Games and Culture, vol. 8, no. 4, pp. 186–214, 2013.

[8] A. Steinle, C. Rüede, and K. Reiss, "Validity of multiple-choice digital formative assessment for assessing students' (mis)conceptions: Evidence from a mixed-method study in algebra," ZDM – Mathematics Education, vol. 56, pp. 713–726, 2024.

[9] A. Churches, "Bloom's taxonomy blooms digitally," Tech & Learning, 2008. [Online]. Available: https://www.techlearning.com.

[10] G. J. Hwang, P. H. Wu, and C. C. Chen, "An online game approach for improving students' learning performance in web-based problem-solving activities," Computers & Education, vol. 59, no. 4, pp. 1246–1256, 2015.

[11] R. Luckin, W. Holmes, M. Griffiths, and L. B. Forcier, Intelligence unleashed: An argument for AI in education. Pearson, 2016.

[12] J. Nielsen, Usability engineering. Morgan Kaufmann, 1994.

[13] J. Lazar, J. H. Feng, and H. Hochheiser, Research Methods in Human–Computer Interaction, 2nd ed. Waltham, MA, USA: Morgan Kaufmann, 2017.

[14] S. Wang, Z. Liu, and Y. Li, "Artificial intelligence in education: A systematic literature review," Expert Systems with Applications, vol. 238, p. 121979, 2024.

[15] J. L. Plass, R. E. Mayer, and B. D. Homer, "Artificial intelligence in education: A cognitive and learning sciences perspective," Educational Psychologist, vol. 60, no. 1, pp. 1–17, 2025.

[16] J. Garzón, S. Baldiris, J. Acevedo, J. Pavón, and R. Baldiris, "Artificial intelligence in education: Trends, benefits, and challenges—A systematic review," Education Sciences, vol. 15, no. 8, p. 84, 2025.

[17] B. Williamson and R. Eynon, "Historical threads, missing links, and future directions in AI in education," Learning, Media and Technology, vol. 45, no. 3, pp. 223–235, 2020.

[18] C. Zhao, X. Li, and G. Chen, "AI-assisted assessment in higher education: A systematic review," International Journal of Educational Technology in Higher Education, vol. 21, no. 1, p. 39, 2024.

[19] M. Sailer and L. Homner, "The gamification of learning: A meta-analysis," Educational Psychology Review, vol. 32, no. 1, pp. 77–112, 2020.

[20] X. Ni, Z. Nuryana, S. Lu, and W. Xu, "A systematic literature review of mathematics interactive games in elementary education," Interactive Learning Environments, pp. 1–28, 2025.

[21] Y. Pan, F. Ke, and X. Xu, "A systematic review of the role of learning games in K–12 mathematics education," Educational Research Review, vol. 36, p. 100428, 2022.

[22] J. Mayrhofer, B. Rengs, and A. Eder, "Current state of gamification design in secondary mathematics education: A systematic literature review," Education Sciences, vol. 15, no. 2, p. 215, 2025.

[23] E. Susantini, Y. M. Sari, M. I. Marzuqi, and P. V. Asteria, "EduQuestioning mobile learning application: A catalyst for developing HOTS-based assessment questions referring to Revised Bloom's Taxonomy," Research & Practice in Technology Enhanced Learning, vol. 20, 2025.

[24] H. N. Boone and D. A. Boone, "Analyzing Likert data," Journal of Extension, vol. 50, no. 2, pp. 1–5, 2012.

[25] M. Milosz, M. Plechawska-Wójcik, and M. Dzieńkowski, "Testing the quality of the mobile application interface using various methods—A case study of the T1DCoach application," Applied Sciences, vol. 14, no. 15, p. 6583, 2024.

[26] V. J. Shute and D. Zapata-Rivera, "Adaptive educational systems," in Adaptive technologies for training and education, P. Durlach and A. Lesgold, Eds. Cambridge University Press, 2012, pp. 7–27.
[27] B. Maqbool, A. Alsaedi, and A. Ahmad, "Potential effectiveness and efficiency issues in usability evaluation methods: A systematic review," Journal of Systems and Software, vol. 207, p. 111858, 2024.
[28] A. Field, Discovering Statistics Using IBM SPSS Statistics, 5th ed. London, U.K.: Sage, 2018.

AUTHORS

Dr. Siti Sabariah Abas is a lecturer at the Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), Terengganu, Malaysia. Her academic specialisation is in Numerical Analysis.

Dr. Nazirah Abd. Hamid is a lecturer at the Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), Terengganu, Malaysia. Her academic specialisation is in Information Security.

Dr. Wan Malini Wan Isa is a lecturer from the Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), Terengganu, Malaysia. Her academic specialisation is in Serious Games.

Assoc. Prof. Dr. Mumtazimah Mohamad is a lecturer from the Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), Terengganu, Malaysia. Her academic specialisation is in Machine Learning.

Maizan Mat Amin is a lecturer from the Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin (UniSZA), Terengganu, Malaysia. Her academic specialisation is in User Interface and User Experience.