Fitrah: Journal of Islamic Education
Vol 6 No. 1 (2025)
Available online at http://jurnal.staisumatera-medan.ac.id/fitrah

P-ISSN: 2723-3847
E-ISSN: 2723-388X

DEVELOPMENT OF A DIGITAL ASSESSMENT MODEL FOR AUTOMATED SHORT ESSAY SCORING IN ISLAMIC EDUCATION STUDY PROGRAM

Evi Susilawati[1], Epi Supriani Siregar[2], Susi Ekalestari[3], Evi Syuriani Harahap[4]
[1,3]Universitas Islam Sumatera Utara, [2]Universitas Pembinaan Masyarakat Indonesia, [4]Universitas Negeri Medan
E-mail: evisusilawati@fkip.uisu.ac.id, episupsiregar1216@gmail.com, susi.ekalestari@sastra.uisu.ac.id, eviharahap21@unimed.ac.id

How to Cite:
Susilawati, E., Siregar, E.S., Ekalestari, S., Harahap, E.S. (2025). Development of a Digital Assessment Model for Automated Short Essay Scoring in Islamic Education Study Program. *Fitrah: Journal of Islamic Education*, *6*(1), 208-223. https://doi.org/10.53802/fitrah.v6i1.1195.

KEYWORDS:
Automated Short Essay Scoring, Digital Assessment, Islamic Education

ABSTRACT
Advances in digital technology have driven the transformation of assessment systems in education, including in the Islamic Education Study Program at the Universitas Islam Sumatera Utara. The manual essay assessment process is often considered inefficient, time-consuming, and prone to subjectivity on the part of assessors. This study aims to develop a Digital Automated Short Essay Scoring Assessment Model designed to enhance the effectiveness, speed, and objectivity of essay assessment for students. This study employs a research and development approach with a four-stage model, namely the definition stage, design stage, development stage, and dissemination stage, with a focus on the first three stages. Research data were collected through interviews, observations, document reviews, digital-based essay tests, and user satisfaction questionnaires. Data analysis was conducted using qualitative and quantitative descriptive approaches. The results of the study indicate that the Automated Short Essay Scoring Digital Assessment Model is valid according to experts, feasible for use, and capable of improving the efficiency of the assessment process and student motivation. This model has positive implications by supporting adaptive learning, facilitating instructors in providing quick, accurate, and structured feedback, and encouraging improved quality of academic interaction between instructors and students.

KATA KUNCI:
Automated Short Essay Scoring, Penilaian Digital, Pendidikan Islam

ABSTRAK
Kemajuan teknologi digital telah mendorong transformasi sistem penilaian dalam dunia pendidikan, termasuk pada Program Studi Pendidikan Agama Islam di Universitas Islam Sumatera Utara. Proses penilaian esai yang dilakukan secara manual sering dinilai kurang efisien, memakan banyak waktu, dan rentan terhadap subjektivitas penilai. Penelitian ini bertujuan untuk mengembangkan Model Asesmen Digital Automated Short Essay Scoring yang dirancang untuk meningkatkan efektivitas, kecepatan, dan objektivitas dalam penilaian esai mahasiswa. Penelitian ini menggunakan pendekatan penelitian dan pengembangan dengan model empat tahap, yaitu tahap pendefinisian, tahap perancangan, tahap pengembangan, dan tahap penyebaran, dengan fokus pada tiga tahap pertama. Data penelitian dikumpulkan melalui wawancara, observasi, telaah dokumen, tes esai berbasis digital, serta kuesioner kepuasan pengguna. Analisis data dilakukan dengan pendekatan deskriptif kualitatif dan kuantitatif. Hasil penelitian menunjukkan bahwa Model Asesmen Digital Automated Short Essay Scoring dinyatakan valid oleh para ahli, layak digunakan, serta mampu meningkatkan efisiensi proses penilaian dan motivasi belajar mahasiswa. Model ini memiliki implikasi positif dengan mendukung pembelajaran adaptif, mempermudah dosen dalam memberikan umpan balik cepat, akurat, terstruktur, serta mendorong peningkatan kualitas interaksi akademik antara dosen dan mahasiswa.

INTRODUCTION

The rapid development of digital technology has revolutionized almost all aspects of education delivery, from curriculum design to evaluation of learning outcomes (Amalia et al., 2025; Zahraini et al., 2025). In the field of Islamic Education Study Program, this transformation offers significant opportunities to improve access, relevance, and quality of learning (Naimi et al., 2025; Zein, 2024). Today, Islamic Religious Education learning activities demand innovation to ensure that Islamic values are maintained amidst the pressures of accelerating technology (Ballianie et al., 2024; Fauzi et al., 2025). Recent studies on Islamic Religious Education program reform emphasize the need for integrative strategies that can bridge digital advances with distinctive spiritual and ethical values (Rahma et al., 2024).

At the level of assessment models, there has been a shift from paper-based tests to online assessments, reducing reliance on manual procedures that are slow and prone to bias. Research on the adoption of digital assessment applications in the Islamic Education Study Program shows an increase in efficiency, although its use is still hampered by the limited digital literacy of lecturers and the availability of infrastructure (Fauzi et al., 2025). In this context, previous research highlights the urgent need for a digital assessment model that is not only fast and objective, but also adaptive to the specific characteristics of the Islamic Education Study Program at the Universitas Islam Sumatera Utara.

Automated Short Essay Scoring comes as an artificial intelligence-based solution to score essay answers automatically (Susilawati, Khaira, Rambe, et al., 2023; Susilawati, Lubis, et al., 2023). In the past decade, research on Automated Short Essay Scoring has evolved from statistical feature-based techniques to advanced neural network models, which is now one of the most prominent applied areas of natural language processing in education (Susilawati, Lubis, et al., 2023). However, most research still focuses on improving assessment accuracy on a standard international corpus and often neglects aspects of contextualization as well as feedback functionality.

The latest wave of research adds a layer of generative intelligence. The development of the Automated Short Essay Scoring model was able to provide richer written feedback and demonstrated up to eighty-seven percent accuracy on a corpus of essay answers, outperforming the consistency of human judgments (Susilawati, Khaira, Ali, et al., 2023). At the national level, research on Automated Short Essay Scoring is still relatively new. The transformation of education in the digital era demands an assessment system that is more efficient, fair, and able to provide quick feedback (Susilawati, Khaira, Rambe, et al., 2023). The current assessment model has not taken into account the uniqueness of local content and religious values contained in the Islamic Education Study Program.

This causes a mismatch between the evaluation indicators in the generic Automated Short Essay Scoring and the learning outcomes in

the Islamic Education Study Program, such as reasoning based on *naqli* arguments (scriptural evidence), contextual relevance to society, and spiritual and moral values. Without adjustment, the assessment results have the potential to be biased and not reflect the holistic ability of students in the Islamic scientific domain. Therefore, the development of a contextual and relevant digital Automated Short Essay Scoring assessment model is an urgent need to support the transformation of authentic assessment in the Islamic Education Study Program.

In the context of curriculum policy in the Islamic Education Study Program, students are given the freedom to study across study programs and campuses (Nur Hadi et al., 2024). This scheme indirectly increases the diversity of tasks and assessment models, especially essay-based assessments that require reflective reasoning, critical analysis, and scientific argumentation. To maintain the quality and objectivity of assessment on a large scale as well as the efficiency of lecturer evaluation, an Automated Short Essay Scoring model is needed that not only performs automatic assessment but is also adaptive to variations in learning outcomes (Susilawati, Khaira, Rambe, et al., 2023). Without evaluation innovation, curriculum implementation in the Islamic Education Study Program risks losing academic accountability and speed of learning services.

As an Islamic educational institution, Universitas Islam Sumatera Utara has a mission not only to produce graduates who excel academically, but also spiritually strong and characterized. Therefore, the assessment system developed must be able to capture character values in student papers, such as honesty, responsibility, spiritual depth, and academic ethics. Conventional systems only assess the linguistic dimension or argument structure. The digital Automated Short Essay Scoring model designed in this study will integrate parameters that evaluate Islamic values in student narratives (Brew & Leacock, 2013; Ramesh & Sanampudi, 2022).

This research has a clear distinction compared to previous studies on Automated Short Essay Scoring. The study conducted by Amalia et al. (2025) developed automatic short answer scoring based on Bidirectional Encoder Representations from Transformers for Indonesian language with a focus on reducing bias and improving scoring consistency. M eanwhile, research by Susilawati, Lubis, et al.(2023) focuses the development of Automated Short Essay Scoring on measuring the character of students through digital technology, but has not contextualized this system in depth with the religious curriculum. Research conducted by Joseph et al.,(2024) highlights strategies for implementing digital transformation in education, but does not specifically address the need for spiritual value-based evaluation. Zahraini et al. (2025) emphasized the importance of Islamic education reform in the digital era, but did not integrate artificial intelligence technology in the essay grading system. The uniqueness of this research lies in the development of digital Automated Short Essay Scoring that is tailored to

the curriculum, learning outcomes, and indicators of religious competence in the Islamic Education Study Program. This model incorporates spiritual, ethical, and social relevance dimensions that are rarely explored in previous research. Thus, this research makes an original contribution in connecting artificial intelligence technology with the authentic evaluation needs unique to Islamic education in Indonesia.

Based on the explanation above, this research aims to develop a valid, reliable, and effective digital Automated Short Essay Scoring assessment model in the Islamic Education Study Program at the Universitas Islam Sumatera Utara. The research focus is directed at designing an artificial intelligence-based essay scoring system that is not only able to assess objectively and consistently, but also relevant to the characteristics of the Islamic Religious Education curriculum, including reasoning based on *naqli* postulates, spiritual values, and scientific ethics. The contribution of this research lies in the effort to present a digital assessment model that is contextualized with the needs of Islamic learning in the era of digital transformation. This research is expected to provide innovation in the learning evaluation system that is faster, more accurate, and provides constructive feedback for lecturers and students. In addition, the results of this study can be a reference for the development of digital assessments in other fields of religious education.

THEORY DESCRIPTION

The development of digital technology has brought significant impacts in the field of education, including in the evaluation of learning outcomes (Uto, 2021; Zein, 2024). Traditional paper-based assessment systems are considered less efficient, time-consuming, and prone to subjectivity. In this context, various efforts have emerged to integrate artificial intelligence technology in assessment, one of which is through the development of Automated Short Essay Scoring (Beseiso et al., 2021; Rahma et al., 2024). This model serves to score essay answers automatically by utilizing natural language processing and machine learning techniques. According to Brew and Leacock (Brew & Leacock, 2013), automated scoring not only speeds up the evaluation process, but can also improve scoring consistency over manual scoring.

In the last decade, Automated Short Essay Scoring evolved from statistical feature-based methods towards advanced neural network models. Ramesh and Sanampudi (Ramesh & Sanampudi, 2022) explain that the latest generative models can analyze text meaning more contextually, while providing richer and deeper written feedback. These advantages become relevant for modern learning that demands personalization, speed, and accuracy in providing assessment.

In the context of religious education, the need for automated assessment becomes more complex as it must take into account distinctive spiritual, moral and scientific values. Generic assessment systems often fail to assess aspects of reasoning based on naqli arguments (evidence from scripture), social relevance, and spiritual

depth of students. Therefore, the development of an Automated Short Essay Scoring model tailored to the Islamic Religious Education curriculum is imperative. Contextualized assessment models not only serve to measure cognitive abilities, but also assess affective aspects and academic ethics, such as honesty and responsibility (Erawati & Adnyana, 2024; Zahraini et al., 2025).

Theoretical studies on the integration of artificial intelligence in education also highlight the need for system adaptation to local conditions. Amalia, Lydia, Muchtar, Manik, Sinu, and Gunawan (2025) emphasized the importance of bias reduction in automated assessment through algorithms based on Bidirectional Encoder Representations from Transformers. Meanwhile, research (Susilawati, Khaira, Ali, et al., 2023) shows that although digital technology can accurately assess linguistic aspects, this system has not been able to measure religious values in depth without setting relevant parameters. With reference to these studies, this research fills the gap by developing an Automated Short Essay Scoring model that combines artificial intelligence technology with spiritual values, ethics, and social relevance in accordance with the characteristics of the Islamic Education Study Program curriculum. This approach is expected to produce a more comprehensive and authentic essay assessment (Rana et al., 2023).

The framework of this research departs from the problem of low efficiency and objectivity of paper-based essay assessment in the Islamic Education Study Program. Digital transformation encourages the development of artificial intelligence-based assessment systems that are faster, more accurate, and consistent, one of which is through Automated Short Essay Scoring. A research and development approach is used to design a contextualized digital assessment model, by integrating spiritual values, ethics, and social relevance in the assessment parameters. This model is expected to be an authentic assessment innovation that supports the quality of Islamic learning in the digital era.

METHOD

This research uses a Research and Development approach to develop a Digital Automated Short Essay Scoring Assessment Model in the Islamic Education Study Program at the Faculty of Islamic Studies, Universitas Islam Sumatera Utara. The goal is to improve the efficiency and objectivity of essay assessment, as well as support adaptive learning. The 4 D Model development model (Define, Design, Develop, Disseminate) was applied up to the Develop stage.

The Define stage includes needs analysis through interviews with lecturers and students, as well as document review such as syllabus, lesson plans, and previous assessment results to identify challenges in manual assessment. The Design stage includes designing the initial prototype of the Automated Short Essay Scoring assessment model, including the preparation of an Islamic Religious Education value-based assessment rubric that integrates cognitive and affective

aspects, as well as designing an automated technology-based assessment system.

The Develop stage includes model validation by experts, namely learning technology experts from the Islamic Education Study Program and digital assessment system experts. Furthermore, revisions were made based on expert input and limited trials to students of the Islamic Education Study Program at the Faculty of Islam, University of Islam, North Sumatra to assess the feasibility, ease of use, and effectiveness of the digital assessment model. all of these stages are illustrated as follows:

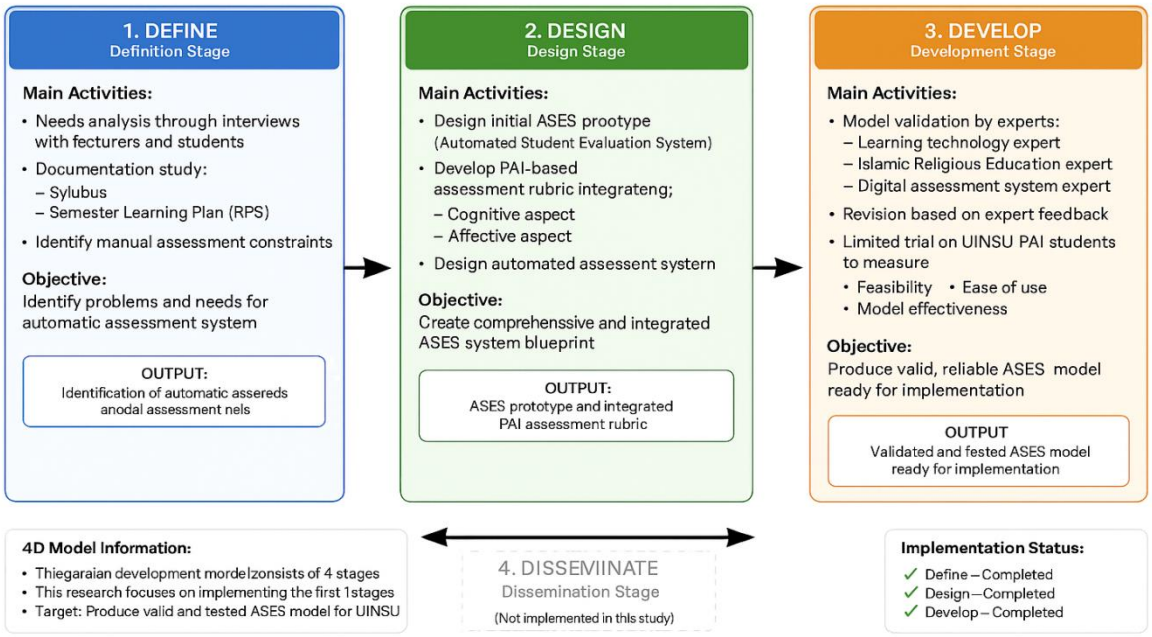

Figure 1. 4 D Model Development and Research

The subjects of this research consisted of 3 lecturers and 30 students of the Islamic Education Study Program at the Faculty of Islamic Studies, Universitas Islam Sumatera Utara. The lecturers acted as the main informants in the interview to identify the obstacles of manual essay assessment, as well as being validators in assessing the feasibility of the assessment rubric and the digital Automated Short Essay Scoring system. Meanwhile, students acted as test users of the digital assessment prototype by working on essay questions based on the Islamic Religious Education curriculum. The results of student answers are used to test the effectiveness and accuracy of the system in scoring essays automatically.

The object of the research is the digital Automated Short Essay Scoring assessment model which includes the development of a technology-based system, an assessment rubric based on Islamic Education values, and a prototype user interface. These three components are designed to combine the assessment of cognitive, affective, and spiritual dimensions in one automated assessment framework.

The research instruments used include interview sheets to explore the experiences and needs of lecturers and students, expert validation

sheets to assess the feasibility of assessment rubrics and digital systems, digital essay questions based on learning outcomes, and user satisfaction questionnaires to measure the experiences of students and lecturers after using the system.

Data collection was carried out through in-depth interviews with lecturers, direct observation of students during the trial process, review of learning documents such as syllabus, lesson plans, and analysis of manual assessment results, and questionnaires filled out by users. Data were analyzed using mixed methods, namely qualitative analysis to interpret the results of interviews and observations, and quantitative analysis with descriptive statistics to calculate the percentage of expert validation and the level of user satisfaction with the digital Automated Short Essay Scoring assessment model.

RESULTS AND DISCUSSION

Analysis of the 4D Model Development Stages

The research results in the Define phase showed an urgent need for the development of a digital Automated Short Essay Scoring assessment model that is more efficient, objective, and in line with Islamic values. The needs analysis obtained through interviews with lecturers and students revealed that manual essay scoring so far requires a long time, is prone to subjectivity, and is less able to provide in-depth and structured feedback. Therefore, the digital assessment model designed should be able to produce a comprehensive evaluation, systematically document the results, and provide constructive feedback. The findings of the Define phase analysis are summarized in the following table.

Table 1. Results of the Define Phase Analysis

No.	Source of Findings	Main Issues Identified	Implications for Automated Short Essay Scoring Model Development
1	Interview with lecturers	Essay assessment takes a long time, especially in large classes (≥ 40 students)	Automation is needed to shorten grading time and streamline score delivery
2	Interview with students	Scores are perceived as inconsistent; high subjectivity between assessors	Standardized digital rubrics and scoring algorithms are required for objectivity
3	Lecturers and students	Feedback is often brief and lacks depth due to time constraints	The system should auto-generate structured comments covering content, logic, sources
4	Syllabus and Semester Learning Plan review	Islamic values (faith, ethics, academic integrity) are not explicitly integrated	The Automated Short Essay Scoring rubric should include spiritual-evaluative indicators aligned with learning outcomes
5	Review of previous assessments	Assessments were paper-based, varied in format, and not systematically archived.	A centralized digital platform is needed to store scores and feedback systematically

The results of the needs analysis show that the manual essay assessment applied to the Islamic Education Study Program at the Faculty of Islamic Studies, Universitas Islam Sumatera Utara has not been able to provide a fast, consistent, and in-depth evaluation. Lecturers stated that the time needed to check a set of essays is quite long, especially if the number of students reaches more than forty people per class. This condition has an impact on the delay of feedback to students, so that the learning process becomes less responsive to the development of students' abilities. In addition, the manual grading system is often affected by subjectivity, such as differences in perceptions between raters in understanding the quality of arguments, essay structure, and the depth of student analysis.

Students also expressed a desire for a more transparent assessment model that can provide an understanding of the weaknesses and strengths of their answers. Some students felt that the comments or evaluation notes provided by lecturers were often too general and did not help them significantly improve the quality of their essays. Therefore, there is an urgent need for a digital assessment model based on Automated Short Essay Scoring that can provide detailed, structured, and well-documented feedback.

The assessment rubric used in this model must be adapted to the characteristics of Islamic Religious Education learning. The assessment aspects are not only limited to the linguistic structure or logic of the argument, but also include an assessment of the depth of understanding of Islamic values, the ability to link arguments with naqli arguments (evidence from holy books), and the application of academic ethics such as honesty and responsibility. The integration of this spiritual dimension is the main differentiator of the Automated Short Essay Scoring model developed in this study.

The needs table analysis shows that lecturers expect a system that is able to reduce administrative workload, provide speed in evaluation, and still maintain the quality of assessment. Students, on the other hand, want automated feedback that is accurate, clear, and in accordance with applicable academic standards. These findings reinforce the urgency of developing a digital assessment model that is adaptive, objective, and able to assess both cognitive and affective dimensions. The existence of such a model is expected to not only improve learning efficiency, but also provide added value in building students' spiritual character through a more authentic and contextual evaluation with the vision of Islamic education in the digital era.

Subsequent analysis in the Design phase resulted in an initial prototype of an automated essay scoring model that integrates a rubric based on Islamic Education values with an automated digital scoring system utilizing Natural Language Processing. This combination directly addresses the main challenges identified

in the Define phase, such as inefficiency and subjectivity in manual essay scoring. The design also introduces an automated feedback mechanism capable of providing structured and in-depth comments, thus enriching the students' learning experience. In addition, the responsive and user-friendly interface design allows lecturers and students to access the system on various devices, thus increasing the potential for integration within the Islamic Education Study Program at Universitas Islam Sumatera Utara. The results of the analysis at this stage are presented as follows:

Table 2. Analysis Results of the Design Phase

No	Design Component	Key Features	Description / Implementation	Rationale for ASES Development
1	Essay-Scoring Rubric	- Islamic values alignment - Cognitive indicators (concept mastery, argument analysis, depth of analysis) - Affective indicators (integrity, academic honesty)	Rubric developed with Islamic Education lecturers, featuring five performance levels with descriptors and examples from Qur'anic and Prophetic texts	Ensures scoring objectivity and integration of spiritual-ethical outcomes per the Islamic Education curriculum.
2	Digital Automation System	- Web-based interface (responsive) - Basic NLP module (tokenization, word spotting, sentence-length heuristics) - Auto-score generator with manual override option	Built with open-source frameworks (Flask/JavaScript), NLP content filters for learning cues, argument accuracy, and cadence patterns, and then produces an initial score that lecturers can confirm or adjust.	Reduces grading time, flags poorly written essays or learning gaps, and provides a transparent audit trail of score adjustments.
3	Feedback Engine	- Template comments tied to rubric cells - Color-coded highlighting of missing elements	When a score is assigned, the system provides pre-written feedback and highlights relevant parts of the student's text	Delivers faster, clearer feedback, addressing poor content due to lack of depth in the Define stage.
4	User Experience (UX)	- Single sign-on via campus credentials - Dashboard for batch upload/download of essays - Mobile-friendly view	Prototype tested on desktops, tablets, and smartphones; average load time < 2s on campus Wi-Fi.	Guarantees accessibility for both lecturers and students, encouraging easy adoption.

Table 2 illustrates the results of the analysis in the Design phase which contains key components in the development of the Automated Short Essay Scoring model based on Islamic Religious Education values. The first component is an essay scoring rubric designed with the involvement of Islamic Religious Education lecturers to ensure alignment with the curriculum and learning outcomes. The rubric contains five levels of performance with clear

descriptions and relevant examples sourced from the Qur'an and hadith texts. The existence of cognitive indicators, such as concept mastery, argumentation analysis, and depth of analysis, is combined with affective indicators such as integrity and academic honesty. This integration is the basis for ensuring the objectivity of assessment while instilling spiritual-ethical values that are unique to the context of Islamic education.

The second component is a web-based digital automation system with a responsive interface. The system is built using open-source frameworks such as Flask and JavaScript, equipped with natural language processing modules that function for tokenization, keyword tagging, and sentence length analysis. The system automatically generates an initial score that can be confirmed or adjusted by the lecturer. This approach not only speeds up the scoring process, but also creates a transparent audit trail, so that any score adjustments can be tracked to maintain the accuracy and fairness of the evaluation.

The third component is a feedback engine designed to provide automated comments according to the rubric cells. When a score is given, the system automatically generates relevant written feedback as well as marks sections of the student's text that need improvement through colored highlighting. This feature ensures that students receive a quick, clear, and constructive evaluation, thus understanding both the strengths and weaknesses of their essays.

The fourth component is the user experience aspect. The prototype was tested on various devices-desktops, tablets, and smartphones-with an average load time of less than two seconds on the campus network. The integration of a single sign-on feature using campus credentials and a dashboard that allows bulk uploading and downloading of essays were important factors supporting ease of adoption.

The overall analysis shows that the design strikes a balance between speed, objectivity, and Islamic value-based assessment. The system not only improves on the weaknesses of slow and subjective manual methods, but also presents a more comprehensive approach to evaluation by integrating cognitive, affective, and spiritual dimensions.

The activities carried out in the Design phase provided a strong foundation to proceed to the Develop phase, which involved expert validation and limited trials to ensure the feasibility and effectiveness of the ASES model in the real context of the Islamic Education Study Program.

Furthermore, the results of the Develop phase showed that the Automated Short Essay Scoring Model was rated highly valid by the experts, with an average score of 89.6%. The assessment rubric based on Islamic values is considered relevant and applicable, while the digital system is considered strong enough to be

used in the assessment process in the digital ASES model. The description of research activities in the Develop phase is presented in the following table:

Table 3. Analysis Results of the Develop Phase

No	Aspect	Component	Evaluation Result	Remarks/notes
1	Expert Validation	Essay Scoring Rubric	95% of Indicators deemed relevant and contextuality appropriate	Category: High Valid
		ASES Digital System	Considered feasible and functional; AI development suggested for future upgrades	Basic functionality meets current needs; NLP features recommended
		User interface	Rated user friendly and accessible	Responsive on both desktop and mobile devices
		Average Expert Validation Score	89.6 %	Falls into highly valid category
2	Limited trial	Student Feedback (30 students)	86% felt more motivated to write essays	Due to quick and clear feedback
		Lecturer Feedback (3 instructors	90% felt more assisted and efficient in grading compared to manual assessment	Time and objectivity significantly improved
		Time Efficiency	Grading time reduced by up to 60%	Reduced from an average of 3 days to 1 day for mass essay assessment
		Scoring accuracy and consistency	Improved with standardized digital rubric	Enables more consistent scoring across different instructors

The evaluation results of the development of a digital assessment model for Automated Short Essay Scoring in the Islamic Education Study Program showed significant findings, both at the expert validation stage and the limited trial. Expert validation of the essay scoring rubric revealed that 95% of the indicators were considered relevant and appropriate to the learning context. This figure confirms that the rubric structure has been able to integrate the cognitive and affective dimensions expected in the Islamic Religious Education curriculum. The validity category classified as "very valid" strengthens the belief that this rubric can be used as a standard reference for digital-based essay assessment.

The components of the Automated Short Essay Scoring digital assessment system are considered feasible and functional, with a note that further development based on artificial intelligence is recommended to improve the accuracy of natural language processing. The existing Natural Language Processing features have met the basic needs in detecting students' linguistic and argumentation patterns, although the potential for future feature enrichment is great. The user interface is also considered user-friendly and easily accessible through both desktop and smartphone devices. The average expert validation score of 89.6% further strengthens the overall quality of the system design.

A limited trial involving 30 students showed that 86% of them felt more motivated to write essays because of the quick and clear feedback. The automated feedback generated by

the system was able to provide specific clues about the strengths and weaknesses of their writing. From the lecturer's perspective, 90% of respondents felt that the system provided significant support, especially in terms of grading efficiency compared to manual methods. Grading time that usually takes an average of three days for a large class can be cut down to one day with a 60% reduction in time.

Accuracy and consistency of grading improved thanks to the use of standardized digital rubrics. This allows for more uniform assessment results among different lecturers, reducing the potential for subjectivity and score gaps. Overall, these findings indicate that the developed Automated Short Essay Scoring digital assessment model is not only able to improve efficiency, but also answers the needs of authentic evaluation in the Islamic Education Study Program. The integration of Islamic values through a spiritual-ethics-based rubric makes this model relevant to support the vision of Islamic education in the digital era, and has the potential to become an innovative assessment model for similar institutions in Indonesia.

Analysis of Research Findings and Distinctions

The results of this study show that the development of the Digital Assessment Model Automated Short Essay Scoring in the Islamic Education Study Program at the Universitas Islam Sumatera Utara has a significant impact on increasing the effectiveness and efficiency of student essay assessment. The transformation from a manual assessment system to a technology-based system has answered fundamental challenges in the evaluation process, such as limited lecturer time, high subjectivity of assessment, and lack of speed in providing feedback to students. This is in line with the views of Joseph, Onwuzulike, and Shitu (2024) who emphasize that the adoption of technology in educational evaluation can increase objectivity and accelerate academic workflow, while providing a learning experience that is more adaptive and responsive to the needs of learners.

Expert validation of the Automated Short Essay Scoring model showed that 95% of the indicators in the assessment rubric were considered relevant and in accordance with the context of the Islamic Religious Education curriculum. This rubric not only assesses cognitive dimensions such as concept mastery, clarity of argument, and depth of analysis, but also assesses affective aspects such as integrity, academic honesty, and spiritual depth. According to the theory of value-based evaluation proposed by Bloom(1956) in the taxonomy of learning, the ideal assessment should include cognitive, affective, and psychomotor domains. The integration of Islamic values in the Automated Short Essay Scoring rubric makes the essay assessment more holistic, because students are not only measured by their analytical skills, but also by their ability to internalize ethical and moral principles into their writing. This is reinforced by the views of

Susilawati et al. (2023) who emphasized the importance of incorporating spiritual values in evaluation design for Islamic education to be relevant to the goals of character and moral formation.

The limited trial of the Automated Short Essay Scoring model showed a positive response from the 30 students involved. As many as 86% of students felt more motivated to write essays because of the quick, clear, and structured feedback. This is consistent with Deci and Ryan's theory of learning motivation (2012) in Self-Determination Theory, which states that relevant and timely feedback can increase students' intrinsic motivation. Meanwhile, three lecturers involved in the pilot stated that the model was able to reduce the burden of manual grading by 60%, speeding up correction time from an average of three days to just one day for a large number of essays. This efficiency is in line with the research findings of Rahma et al. (2024) which highlighted that the utilization of digital technology in evaluation is able to optimize lecturers' time allocation, so that they can focus more on academic assistance.

In addition, the use of Natural Language Processing technology in Automated Short Essay Scoring allows the system to detect patterns of argumentation, text coherence, and completeness of essay content more objectively. However, the system still provides room for lecturers to do manual override if needed, so that the quality of the assessment is maintained. According to Ramesh and Sanampudi (2022), combining artificial intelligence with human validation is the best approach to ensure a balance between the speed of automation and the accuracy of context-based assessment.

Standardizing assessment through digital rubrics also provides fairness and transparency, as each score is based on clear and documented indicators. This is in line with Black and Wiliam's (2011) view of assessment for learning, where assessment should serve as a tool to promote learning, rather than simply measuring outcomes. Students can understand their weaknesses through systematic feedback, while lecturers can make adjustments to learning strategies based on data obtained from digital assessments.

Overall, the development of the Automated Short Essay Scoring model in the Islamic Education Study Program at the Universitas Islam Sumatera Utara can be said to be an innovative breakthrough that combines technology with Islamic educational values. This model not only strengthens the quality of academic evaluation, but also maintains spiritual relevance in every learning process. This research proves that technological innovation, when contextualized with local needs and religious values, can be a key driver for creating an adaptive, inclusive and character building-oriented education system. Going forward, the development of advanced artificial intelligence-based features and personalization of feedback can enrich the application of Automated Short Essay Scoring, making it a superior digital

assessment model for Islamic education institutions in Indonesia (Carr & Xi, 2010; Shermis & Burstein, 2003; Zimmerman et al., 2018).

The distinction of this research lies in the effort to integrate artificial intelligence technology with an assessment approach based on spiritual and ethical values typical of Islamic Religious Education. Most of the previous research related to Automated Short Essay Scoring focuses on optimizing technical accuracy, scoring efficiency, or reducing rater bias, but does not specifically accommodate the needs of a religious curriculum that is full of value and moral aspects. The research of Amalia et al. (2025) for example, only focuses on developing algorithms based on Bidirectional Encoder Representations from Transformers to overcome assessment bias in a general context, while the research of Joseph, Onwuzulike, and Shitu (2024) highlights digital transformation in education without mentioning the spiritual dimension of assessment. Different from these approaches, this study designed a digital rubric tailored to the learning outcomes of Islamic Religious Education, with indicators that not only assess critical thinking skills and argumentative analysis, but also assess the depth of understanding of *naqli* arguments, social relevance, and moral integrity of the essay writer. This combination of technological aspects and Islamic values has not been explored in previous studies. Therefore, this research makes an original contribution in presenting a digital assessment model that truly fits the characteristics of Islamic education in Indonesia.

CONCLUSION

This study concludes that the development of the Digital Assessment Model Automated Short Essay Scoring in the Islamic Education Study Program at the Universitas Islam Sumatera Utara has successfully made a real contribution in increasing the effectiveness, efficiency, and objectivity of student essay assessment. The transformation from a manual assessment system to a technology-based digital system not only accelerates the evaluation process, but also reduces the level of subjectivity of the assessor. The main findings of this study show that 95% of the assessment rubric indicators were rated as relevant and contextually appropriate by the experts, with the overall validity rate reaching 89.6%. The ASES system proved to be able to reduce assessment time by 60%, as well as provide quick, clear, and structured feedback to students. The novelty of this research lies in the integration of artificial intelligence technology with spiritual, ethical, and moral values that are unique to Islamic Religious Education. The digital rubric designed not only assesses students' analytical skills, but also detects the depth of understanding of naqli arguments, social relevance, and academic integrity. This distinction distinguishes this research from previous studies that only focus on the technical or algorithmic aspects of

assessment. The implication is that this model can serve as a reference for the development of digital assessment systems in other Islamic educational institutions, while strengthening the relevance of religious education to the demands of the digital era. This model also encourages learning that is more constructive, transparent, and oriented towards character building.

REFERENCES

Amalia, A., Lydia, M. S., Muchtar, M. A., Manik, F. Y., & Gunawa, D. (2025). Mitigating Bias and Assessment Inconsistencies with BERT-Based Automated Short Answer Grading for the Indonesian Language. *IAENG International Journal of Computer Science*, *52*(3).

Ballianie, N., Dewi, M., & Fatimah, S. (2024). Inovasi Pembelajaran Pendidikan Agama Islam untuk Generasi Milenial. *Indo-MathEdu Intellectuals Journal*, *5*(6), 8370–8382. https://doi.org/10.54373/imeij.v5i6.2476

Beseiso, M., Alzubi, O. A., & Rashaideh, H. (2021). A Novel Automated Essay Scoring Approach for reliable higher Educational Assessments. *Journal of Computing in Higher Education*, *33*(3), 727–746.

Bloom, B. S. (1956). *Taxonomy of Educational Objectives*. Longmans.

Brew, C., & Leacock, C. (2013). Automated Short Answer Scoring: Principles and Prospects. In *Handbook of automated essay evaluation* (pp. 136–152). Routledge.

Carr, N. T., & Xi, X. (2010). Automated Scoring of Short-Answer Reading Items: Implications for Constructs. *Language Assessment Quarterly*, *7*(3), 205–218. https://doi.org/10.1080/15434300903443958

Deci, E. L., & Ryan, R. M. (2012). Self-Determination Theory. *Handbook of Theories of Social Psychology*, *1*(20), 416–436.

Erawati, N. K., & Adnyana, P. B. (2024). Implementation of Jean Peaget's Theory of Contructivism In Learning: A Literature Review. *Indonesian Journal of Educational Development (IJED)*, *5*(3), 394–401. https://doi.org/10.59672/ijed.v5i3.4148

Fauzi, M. H., Salsabila, S., Diniyati, A. I., Pebriani, A. R., Fithriya, R. A. I., & Suresman, E. (2025). Integrasi Nilai Islam dan Inovasi Pembelajaran dalam Pendidikan Agama Islam di Perguruan Tinggi dalam Perspektif Akademik dan Keagamaan. *Reflection : Islamic Education Journal*, *2*(2), 186–196. https://doi.org/10.61132/reflection.v2i2.771

Joseph, O. B., Onwuzulike, O. C., & Shitu, K. (2024). Digital Transformation in Education: Strategies for Effective Implementation. *World Journal of Advanced Research and Reviews*, *23*(2), 2785–2799. https://doi.org/10.30574/wjarr.2024.23.2.2668

Naimi, N., Nursakinah, N., Sitepu, M. S., & Sitepu, J. M. (2025). Transformasi Pembelajaran Pendidikan Agama Islam di Era Digital: Sebuah Kajian Pustaka. *Waspada (Jurnal Wawasan Pengembangan Pendidikan)*, *13*(1), 105. https://doi.org/10.61689/waspada.v13i1.724

Nur Hadi, M., Supriyatno, T., Marno, & Syaifullah. (2024). Inovasi Pembelajaran Pendidikan Agama Islam Melalui Program Merdeka Belajar Kampus Merdeka Berbasis Outcome Based Education. *Jurnal Mu'allim*, *6*(2), 432–448. https://doi.org/10.35891/muallim.v6i2.5517

Rahma, F., Zain, A., Mustain, Z., & Rokim, R. (2024). Penguatan Nilai-nilai Spiritual dan Moralitas di Era Digital Melalui Pendidikan Agama Islam. *Jemari (Jurnal Edukasi Madrasah Ibtidaiyah)*, *6*(2), 94–103. https://journal.unuha.ac.id/index.php/jemari/article/view/3568

Ramesh, D., & Sanampudi, S. K. (2022). An

Automated Essay Scoring systems: a Systematic Literature Review. *Artificial Intelligence Review*, *55*(3), 2495–2527. https://doi.org/10.1007/s10462-021-10068-2

Rana, J., Daultani, Y., & Goswami, M. (2023). Exploring the Role of Smart Technologies in Supply Chain Digital Transformation. In *Proceedings of the International Conference on Industrial Engineering and Operations Management*. IEOM Society International. https://doi.org/10.46254/an13.20230156

Shermis, M. D., & Burstein, J. C. (Eds. . (2003). *Automated Essay Scoring*. Routledge. https://doi.org/10.4324/9781410606860

Susilawati, E., Khaira, I., Ali, A., Adlina, Z., & Afnaria, A. (2023). Analisis Model Penilaian Berbasis Teknologi Digital untuk Mengukur Perilaku Kejujuran Mahasiswa. *Biblio Couns: Jurnal Kajian Konseling Dan Pendidikan*, *6*(3), 132–138. https://doi.org/10.30596/bibliocouns.v6i3.17062

Susilawati, E., Khaira, I., Rambe, I. H., Sembiring, M. B., Siregar, W., & Octariani, D. (2023). Analysis of Needs for A Digital Automated Short Essay Scoring (Ases) Based Assessment Model In English Learning In Vocational Secondary Schools. *Language Literacy: Journal of Linguistics, Literature, and Language Teaching*, *7*(2), 355–363. https://doi.org/10.30743/ll.v7i2.8248

Susilawati, E., Lubis, H., Kesuma, S., Fitri, H., & Sumantri, P. (2023). Development of an Automated Short Essay Scoring (ASES) Assessment Model as an Assessment Solution for the Digital Era. In *Proceedings of the 8th Annual International Seminar on Transformative Education and Educational Leadership, AISTEEL 2023, 19 September 2023, Medan, North Sumatera Province, Indonesia*. EAI. https://doi.org/10.4108/eai.19-9-2023.2340528

Uto, M. (2021). A Review of Deep-Neural Automated Essay Scoring Models. *Behaviormetrika*, *48*(2), 459–484.

Wiliam, D. (2011). What is Assessment for Learning? *Studies in Educational Evaluation*, *37*(1). https://doi.org/10.1016/j.stueduc.2011.03.001

Zahraini, Z., Akib, A., Rosidin, R., & Sulaeman, O. (2025). Islamic Education Reform in the Digital Age: Challenges and Opportunities for a Modern Curriculum. *Journal of Noesantara Islamic Studies*, *2*(1), 1–11. https://doi.org/10.70177/jnis.v2i1.1841

Zein, M. (2024). Transformasi Pendidikan Islam Di Era Digital, Tantangan dan Solusi Dalam Meningkatkan Kualitas Pembelajaran. *JIPDAS (Jurnal Ilmiah Pendidikan Dasar)*, *2*(3), 146–156. https://ejournal.lpipb.com/backup_ejournal_v1/index.php/jipdas/article/view/434

Zimmerman, W. A., Kang, H. Bin, Kim, K., Gao, M., Johnson, G., Clariana, R., & Zhang, F. (2018). Computer-Automated Approach for Scoring Short Essays in an Introductory Statistics Course. *Journal of Statistics Education*, *26*(1), 40–47. https://doi.org/10.1080/10691898.2018.1443047