Beyond the Euphoria: A Balinese-AI User Interface for Assessing Character, Thought, and Action in Tourism Schools

Gusti Ayu Dessy Sugiharni [1], Putu Herny Susanti [2], Putu Sabda Jayendra [3]

[1] *Department of Digital Business, Institut Pariwisata dan Bisnis International, Jalan Kecak No. 12 Gatot Subroto Timur, Denpasar, Indonesia*
[2] *Department of Tourism Planning and Development, Institut Pariwisata dan Bisnis International, Jalan Kecak No. 12 Gatot Subroto Timur, Denpasar, Indonesia*
[3] *Department of Tourism Management, Institut Pariwisata dan Bisnis International, Jalan Kecak No. 12 Gatot Subroto Timur, Denpasar, Indonesia*

***Abstract* – This study aims to design, develop, and validate a culturally embedded, AI-driven e-assessment user interface (UI) that accurately measures these three competencies while preserving local wisdom. Following the full ten-stage Borg & Gall research and development framework, the research encompassed literature review, needs analysis, initial UI prototyping, and two rigorous empirical testing phases. The Alpha Test involved 2 informatics experts, 2 educational experts, 15 evaluators, and 45 teachers; the Beta Field Trial engaged 2 informatics experts, 2 educational experts, 30 evaluators, and 65 teachers from five tourism vocational high schools in Bali. The data used in this research were gathered from several sources: Tests of functionality, heuristic evaluations, the system usability scale (SUS), the technological acceptance model (TAM) questionnaires, structured observations, and detailed interviews. The Alpha Test results yielded a mean score of 89.23% (excellent), and improvements were made to the AI model's accuracy, the richness of scenarios, and the API's security, as well as the addition of an explainable AI (XAI) reasoning panel.**

The Field Trial results yielded an almost identical mean score of 89.20%, and the TAM scores for perceived usefulness and perceived ease of use were also 89.20%. The validated prototype demonstrates that AI design with a culturally centered, responsive approach will not only deliver superior acceptance and usability but also shift the modernization euphoria paradigm toward culturally responsive design.

***Keywords* – User interface, Artificial Intelligence, e-assessment, Samskara-Cesta-Krtya, Balinese-culture.**

DOI: 10.18421/TEM153-79
https://doi.org/10.18421/TEM153-79

Corresponding author: Gusti Ayu Dessy Sugiharni,
Department of Digital Business, Institut Pariwisata dan Bisnis International, Jalan Kecak No. 12 Gatot Subroto Timur, Denpasar, Indonesia
Email: dessy.sugiharni@ipb-intl.ac.id

Received: 10 September 2025.
Revised: 24 February 2026.
Accepted: 03 March 2026.
Published: 27 August 2026.

1. Introduction

The integration of technology into everyday teaching has become a necessity for enhancing lesson quality and truly understanding what students are learning [1]. This imperative is especially pronounced in vocational schools, such as the tourism vocational high schools scattered across Bali. Today's fast-paced job market calls for assessment methods that can adapt to the times, remain precise, and align with industry demands [2]. As borders blur and technology races ahead, it has been swept into what some call Modernization Euphoria, a moment when every digital trend touches people's lives. The travel business is no exception [3]. While this sense of urgencyopens doors, it also prompts schools to reassess their syllabuses and teaching styles, particularly when it comes to developing character and cultivating the soft skills that students will rely on after graduation [4].

The world of education today stands at the center of this modernization euphoria, where developments in digital technology and Artificial Intelligence (AI) have transformed the paradigms of learning and competency assessment. The Industrial Revolution 4.0 and Society 5.0 demand that educational institutions not only focus on mastery of technical skills (hard skills), but also on character formation, social skills, and creativity that are adaptive to change [5].

Indonesia, particularly Bali as a world tourism destination, relies on vocational high schools to play a strategic role in preparing superior human resources not only in practical abilities but also in preserving local cultural values amid the currents of globalization [6].

Many traditional classroom models still focus on numbers and memorization, paying little attention to how students feel or move [7]. In a tourism setting, where friendliness, a strong work ethic, and quick adaptability are key, the old-school model simply falls short [8]. Those who have undergone vocational training in Bali will interact with tourists from all over the world and, as such, are Bali's frontline workers [9]. To grow into well-rounded individuals, the youth will need three sets of integrated supports, all grounded in traditional Balinese practices: A clear and strong Samskara, a sharp Cesta, and an active Krtya. In simple terms, while Samskara concerns values and character, Cesta pertains to the intellectual and the persuasive, and Krtya is the will to act. The combination of these three hypotheses will make a complete individual [10]. In most cases, these are the easy parts; the hard part is to come up with a practical, appropriate, and reliable way to quantify these elusive, subjective attributes [11].

While widely implemented and accepted in educational institutions, traditional evaluation methods are unable to capture cognitive/ socio-emotional components such as Samskara (character and ethical values), Cesta (social competencies), and Krtya (creativity and innovation). However, these three areas are critically important for tourism vocational students in Bali, as the tourism business demands not only technical skills but also soft skills, culturally appropriate work ethics, and the ability to respond to changes and updates [12]. So far, the evaluation of Samskara, Cesta, and Krtya has relied on the subjective views of the teacher and the interviewer, as well as the formal questionnaire, all of which are easily influenced by bias, unevenness in teacher evaluation, and the limitations of real-time evaluation [13]. Student attitude assessments are another example of teacher evaluation, which by nature is subjective, thereby producing less impartial results [14]. Moreover, in assessing students, existing instruments (attitude assessment sheets and project rubrics) are not capable of comprehensively and holistically capturing student growth and development. Furthermore, traditional measurement tools do not provide instant feedback that can help students improve their competencies independently [15].

Artificial Intelligence is advancing so rapidly that it is beginning to alter the way tests and exams are approached [16]. Because the technology can sift through vast piles of data almost in real time, spot connections humans might easily miss, and adjust its thinking as new information comes in, it opens doors that traditional marketing could not [17]. What is exciting is AI's growing ability to look beyond the numbers and letters on a report card [18]. By applying natural language processing (NLP), sentiment analysis, machine learning, and behavioral analytics, machines are beginning to develop a measurable understanding of intangible qualities such as ethics, empathy, resilience, and creative thought [19]. These qualities, summed up in the ancient terms Samskara, Cesta, and Krtya, have long puzzled educators trying to capture the complete picture of a learner. AI, it seems, may soon provide the language, frameworks, and evidence for those conversations that go well beyond test scores [20].

Machine-learning-powered AI models have the potential to grade essays, oral presentations, and practice simulations with a level of consistency that often surpasses human judges [21]. Virtual Reality (VR) and Mixed Reality (MR) are also emerging as promising tools for simulating real-world scenarios in assessing teamwork, leadership, communication, and emotional regulation [22]. While AI systems have some value for user assessment and feedback, the values and beliefs associated with AI may be challenged and questioned due to the systems' lack of fairness and transparency. Users and stakeholders may believe that AI systems are unfair, favor certain users, and lack transparency in explaining how and why users receive specific feedback and ratings. Furthermore, if the dataset used to train the model reflects outdated stereotypes or is selectively sampled, the results could unfairly skew against certain groups [23]. While AI has immense promise to streamline grading and enhance accuracy, these cracks must first be addressed so that it truly serves all learners [24].

There is no arguing that AI can bring significant benefits to e-assessment, but those advantages will only be realized if the user interface (UI) is well-designed. It does not matter how clever the technology is; if students and teachers cannot figure out how to use it, the whole system falls flat [25]. The right design streamlines every click, cuts down the time people need to learn the ropes, and ultimately leaves users feeling satisfied rather than frustrated [26]. Beyond that, the interface should break down complex data from multiple measures so that everyone, from a tech-savvy student to an anxious teacher, can grasp the results at a glance [27].

Any new app or platform aimed at tourism vocational high school students in Bali needs to reflect who those learners are. That means considering what they like to see on a screen, how comfortable they are with various digital tools, and the traditions they observe every day [28]. Dropping in a Balinese motif here or picking a color palette that echoes local festivals there can help the interface feel like home and catch their attention. Language, layout, and symbols are strongly influenced by culture. Integration of traditional Balinese elements, such as the Tri Hita Karana philosophy (harmony with God, humans, and nature), with modern technology can support cultural preservation [29]. It is essential to address AI biases that often reflect Western ideologies [30].

Like many vocational students in Indonesia, learners from Bali's tourism vocational high schools often struggle to acquire non-technical skills. Aspects such as communication, teamwork, and critical thinking frequently receive less attention than hotel management or travel planning, as shown [31]. Students also miss opportunities to practice and use English in real-world contexts because, in most cases, they do not get to speak with or interact with native English speakers. Moreover, the assigned teaching and learning materials do not focus on or relate to students' contexts and situations, which, in turn, leads to a lack of motivation to learn [32]. Another layer of complexity in this problem is the issue of digital literacy. The students in Bali walk into the classroom with varying levels of comfort and familiarity with cell phones, applications, and various online tools. While some students are very adept at using social media applications such as Instagram and online gaming, others are unable to send an email to submit an assignment. This gap in skill and comfort level puts more stress on teachers [33]. The Euphoria of Modernization offers opportunities such as online booking and virtual tours, but it also poses the challenge of devaluing older ideals. Family and teachers express concerns about the lack of civility and responsibility [34].

To clarify the goals of this research, a practical needs assessment was performed, incorporating informal interviews, brief surveys, and observations of classroom activities. Students, teachers, and employers at local hotels and travel agencies who recruit graduates were consulted. The objectives were as follows: First, to determine, among the students, which non-technical competencies they find most challenging; second, to understand the criteria set by schools to evaluate those skill sets and identify where the evaluation system fails; third, to gauge the level of digital, especially AI testing, and the user interface of the application designed to meet the criteria of being Balinese.

With this in mind, the current study aims to develop a culturally integrated, AI-based e-assessment user interface to measure Samskara, Cesta, and Krtya among students at Bali tourism vocational high school in the context of Modernization Euphoria. By following all ten steps of the Borg & Gall Research and Development framework, this study includes a literature review, a needs assessment, the first UI prototype, and two extensive empirical tests: An initial Alpha Test (trial) with experts and teachers, and a Beta Field Trial with real users in authentic classroom conditions. This study aims to document the entire design process and empirical evidence of usability testing, the AAI (Artificial Archival Intelligence) accuracy assessment, and multiple modifications that culminated in a final, verified prototype. This research outlines several areas in which its findings can be applied. The research can affect how Bali's tourism sector assesses student performance, helping them understand the sector's performance requirements. The research guides schools in fostering the values of Cesta and Krtya, as evidenced in graduates' work ethic and the requisite skills for in-demand jobs. It is further demonstrated through the integration of Artificial Intelligence into the e-assessment process, which, while applicable to the current project, provides a guide for others to develop e-assessment frameworks that are flexible and intelligent to accommodate the rapid changes brought by various forms of modernisation. This research provides guidance for online learning and assessment tool designers on developing simple yet sophisticated assessment systems capable of evaluating multiple, interrelated, complex skills and competencies. These improvements are significant because they equip the education system to address the ongoing changes and the need for cultural sustainability by preparing students to meet the challenges of learning and work, as well as the changing world.

2. Methodology

This investigation adhered to the Borg and Gall Research and Development (R&D) model, which is commendable for the systematic approach to the design and evaluation of instructional materials. Given the dimensions of this study, which included the design of an AI system, culturally responsive instructional materials, and the sophisticated measurement of student competencies, the systematic approach was highly appropriate. All 10 stages of the Borg and Gall model were included in the research. Still, this article will focus only on the cycles of empirical validation for the initial and final products.

The model's stages are research and information collection, design, development of product prototype, alpha testing, product revision, beta testing, approximation of the product revision, field testing of the product, revising the product, and dissemination and implementation. This article focuses on the 6 stages, from the 1st to the 6th, which lead to a validated prototype for wider implementation.

2.1. Research and Information Gathering (Stage 1)

The first phase laid the foundation for all future choices and actions. To ensure that their future prototypes resonate with and address the needs of students from the tourism vocational high schools throughout Bali, the team sought to reinforce the e-assessment interface with well-grounded local realities. The inquiry, therefore, began with an exhaustive review of the existing scholarship across several interconnected fields. The scope of the literature review focused on the intersection of AI and educational assessment, particularly assessment based on efficiency, scale, and customization; the assessment of soft skills using natural language processing, AI, machine learning, behavioral analytics, as well as Virtual Reality and Mixed Reality; and the assessment of soft skills. The frameworks of the Balinese Tri Hita Karana philosophy were applied to the mental impressions, awareness (conscious), and deliberate (purposeful) actions of Samskara, Cesta, and Krtya. The fundamental principles of UI/UX design (such as user control and consistency, error prevention, recognition rather than recall, flexibility and minimalism, and the design of interactive visualizations, dashboards, and filters) were examined, as were the principles of progressive disclosure for visualizing complex data.

The research also looked at the impact of culture on UI design, such as the use of language, design, icons, and the use of traditional Balinese elements, as well as the contextual profile of Indonesian vocational students, such as difficulties in the acquisition of non-technical skills, low motivation, a passive learning tendency, and differences in the level of digital literacy. The "Euphoria of Modernization" concept was discussed as describing the phenomenon of uncritical adoption and the risks of cultural and social erosion.

Alongside the literature review, practical needs analysis involved semi-structured interviews with 45 teachers from five tourism vocational school in Bali, focus groups with 120 students, consultations with 30 industry stakeholders from the hotel, travel agency, and tourism association industries, observation of classroom assessment practice, and surveys on digital literacy and preferences in UI design.

The needs analysis identified the assessment of soft skills, particularly communication, collaboration, problem-solving, and ethical reasoning, as an area that requires greater attention due to the lack of clarity regarding the sophistication of the assessment and the traits being targeted. Considering the range of digital literacy and the expressed need for a culturally tailored design, an adaptive UI was necessary. Respondents voiced concerns around AI, data privacy, and bias. In the context analysis, the Tri Hita Karana philosophy (three causes of prosperity) was identified as the main cultural basis for understanding and framing the measures of Samskara, Cesta, and Krtya, ensuring that evaluative tools are aligned with local culture. To remain relevant, the complexities of Bali's tourism in the aftermath of COVID-19 and the disruptions caused by digital nomad tourism were also analyzed.

2.2. Planning (Stage 2)

The planning phase translated collected data into a development roadmap. Specific learning outcomes for the e-assessment focused on the measurement and development of the Samskara, Cesta, and Krtya capabilities. It required translating philosophical theories into examinable, measurable behaviors within the tourism sector. For instance, the measurement of Samskara was operationalized to focus on ethical disposition, cultural empathy and values, and irresponsible hospitality behaviors within the tourism industry. The measurement of Cesta focused on the absence of critical thinking regarding problem- and decision-making, as well as the presence of strong analytical reasoning in tourism-related case studies. The measurement of Krtya was defined and focused on the absence of practical, collaborative, and adaptable behaviors, as well as the presence of purposeful action within work simulations. In light of these defined behaviors, it was envisaged that the AI technology, given its capacity, would provide accurate behavioral analysis, including through text, voice, video, and simulations of behavioral activities.

The system specification and functionality section outline the detailed planning that went into the system's identification and the design of the UI, AI, technical systems, and ethics. The system's UI is designed to include interactive dashboards for visualizing data from assessments, adaptive assessments, and personal feedback, with progressive disclosure to lower cognitive load and easy navigation. AI enabled the use of natural language processing for analyzing text and voice responses for Samskara and Cesta, machine learning for behavioral pattern analysis in simulations for Krtya, computer vision for the analysis of expressions and gestures in VR situations, and explainable AI to justify/scaffold the scores given.

The technical system and design helped outline the development environment, including the Laravel PHP framework, Python for AI, JavaScript for the interactive UI, and MySQL for the database. System ethics in the design to address bias in AI through culturally relevant training sets, the encryption of student data, transparency in the AI assessment process through XAI, and human oversight for all assessment decisions. The initial validation strategy was outlined, including expert review protocols for content, learning design, and media, which included experts from education, AI, UI/UX, and Balinese culture to ensure relevance.

2.3. Development of Initial Product Form (Stage 3)

At this point, elements from previous stages were used within the initial designs and prototypes. It was achieved through wireframing, information architecture, culturally sensitive visual design, and interaction design. For the wireframes and mockups, at least one was made for the primary screens of the e-assessment system. Screens included the login screen, student and teacher dashboards, assessment interfaces for simulations, essays, and responses, audio recordings, and feedback screens. The information architecture was built to provide easy, quick access to the assessment modules and student feedback. There is a culturally sensitive visual design for Balinese culture. It involves the use of ethnically appropriate and culturally sensitive designs, color combinations, and specific typography, icons, and design arrangements. These made the e-assessment system more culturally sensitive by relaxing its streamlined Western design through traditional Balinese decoration. Interaction design is about the behavioral aspects of the system (including the design of activities, behavioral animations, interfaces, and elements, and ease of use and engagement), while adhering to the principles of Samskara, Cesta, and Krtya. It is done by using interactive simulations to facilitate AI to assess the students.

The login page prioritizes user experience, safety, and brand recognition by incorporating Balinese elements to enhance familiarity and reduce cognitive load. The teacher dashboard is a command center that aggregates advanced visual analytics of students' performances across disciplines of Samskara, Cesta, and Krtya. Teachers can view performance summaries of entire classes and individual students, analyze developments over time via line graphs, interactively filter data by class, term, or competency, access student profiles with XAI-generated feedback, and note qualitative observations. Suitable graph types were utilized: Line graphs to visualize trends, bar graphs for comparisons between students or classes, and radar graphs to display competency profiles across the three disciplines of Samskara, Cesta, and Krtya.

The student dashboard was built on a student-centered AI design philosophy and includes a personal learning map, a skill radar chart of Samskara, Cesta, and Krtya, cultural badge system for achievement motivation, reflective practice via daily dharma tasks, AI personal feedback with explicit and understandable narrative feedback, learning suggestions via AI assessments, goal-setting and tracking, and streamlined access to assessments.

Integration of AI components with the user interface (UI) has enabled seamless integration of AI features. This includes specifying the format of the input data for the AI (i.e., text from essay submissions, audio from verbal response submissions, and behavioral data from simulation activities) and how the AI will analyze this data, how the AI will visualize the output in the user interface, and the user interface design. This design allows for AI-generated feedback to be presented to the user interface in a format that is easy to comprehend and actionable for users, in this case, educators and students, to aid users in recognizing the strengths and weaknesses of the feedback in relation to the Samskara, Cesta, and Krtya feedback. It was followed by developing the initial assessment content by designing items/scenarios that measure the Samskara, Cesta, and Krtya activities, allowing AI to analyze students' relevant dispositions, thoughts, and actions. It was contextualized in the tourism industry in Bali and the Euphoria of Modernization challenge faced by students. Interactive prototypes of the user interface design were developed to simulate the user experience for initial testing and to gather feedback from potential users and experts.

2.4. Initial Trial (Alpha Test) – Stage 4

The Alpha Test was conducted on the AI-based e-assessment model prototype to evaluate the fundamental processes, practical implementation, and other factors related to its usability. As part of this preliminary round, the prototype's functionality was reviewed by experts to assist with debugging and refining the design, as well as to provide feedback on prototype testing outside the organization. The participants included 2 experts in informatics, 2 experts in education, 15 evaluators of education specializing in educational assessment, and 45 teachers from Bali's tourism SMKs, for a total of 62 respondents. Evaluation methodologies consisted of functionality testing to examine all primary features like video upload, NLP analysis, simulations of Krtya, dashboard interactions, and the generation of feedback; heuristic evaluation to apply feedback on the user-interface to inform design based on Nielsen's principles of usability to include visibility, matching the real world, and control of the user, consistency, prevention of errors and recognition, flexibility, and aesthetics; think-aloud protocols in which the experts vocalized their thoughts while interacting with the system to explore usability issues and cognitive friction points; and a system usability scale (SUS) survey which uses a 15-item questionnaire designed for the e-assessment context on a 5-point Likert scale from strongly disagree to strongly agree [35].

The SUS instrument measured perceived usability, learnability, satisfaction, cultural comfort, AI feedback clarity, dashboard clarity, trust in results, and willingness to recommend. SUS scores were calculated using a standard formula where odd-numbered items were scored as scale position minus one, even-numbered items were scored as five minus scale position, then summed and multiplied by 2.5 to obtain a percentage, with mean scores categorized as excellent above 85 percent, good between 70 and 85 percent, fair between 50 and 69 percent, and poor below 50 percent.

2.5. First Revision (Stage 5)

The Cesta motion analysis and Samskara NLP interpretations were enhanced with improved AI model accuracy. Edits were made to enrich further and expand the range of Krtya evaluations. Improvements were made to the Ketya API security and to the teacher dashboard's Artificial Intelligence transparency features. Edits were made to standardize user interface components to improve usability. Changes were made based on specific recommendations of expert reviewers and considering the identified weaknesses before beginning further external trials.

2.6. Field Trial (Beta Test) – Stage 6

Before final implementation, issues arising in real-world use cases need to be identified and resolved. In Bali, the Beta Test included 99 participants. There were 30 evaluators, 65 teachers, 2 informatics specialists, and 2 educational specialists. Each of Bali's five secondary tourism schools had one preset field-trial configuration in locations that exhibited geographic and demographic diversity, including Denpasar, Badung, Gianyar, Tabanan, and Buleleng. These schools also had varying degrees of technological infrastructure and levels of digital literacy.

Using a 5-point Likert scale, participants in the test were given a questionnaire about the Technology Acceptance Model in the fields of perceived usefulness and perceived ease of use, with 8 and 7 items, respectively. The system usability scale, as used in the Alpha Test, was given to all 99 participants along with the 15 instruments used in the previous test.

There was also a participatory observation component during which the research team examined the teacher's use of the system in a real classroom, describing their challenges, methods, and unprompted responses to the system. In the end, semi-structured interviews were conducted with 20 experts, evaluators, and teachers to document the user's experience, the culture of the user's experience, the user's perception of the AI's accuracy, and suggestions for the user's experience.

Quantitative data were drawn from the SUS and the TAM, and descriptive statistics (means and percentages) were subsequently applied, with the data categorized according to descriptive statistics benchmarks. Qualitative data obtained from observations and interviews were analyzed using a thematic framework, in which themes were derived through iterative coding of transcripts and field notes. Triangulation was used to cross-verify findings across different data sources and methods to improve overall trustworthiness.

2.7. Final Revision (Stage 6 – continued)

Some improvements have been made based on Beta Test findings and user suggestions. The main focus was on improving system performance in low-bandwidth environments and on low-spec devices via lazy loading, video compression, and offline mode functionality. Other areas of improvement included: Enhancing UI responsiveness/ease and simplifying XAI explanations for non-technical users; increasing AI accuracy by retraining on Balinese informal language, spell-check variations, and recording guide conditions; expanding the Krtya scenario bank to over 50 branching scenarios; providing users personalized feedback through AI-generated suggestions for improvement; and adding "Report Inaccuracy" functionality for users to report feedback on AI assessments. A closed feedback loop enables continuous improvement to the system.

3. Results

The results are organized according to the empirical testing stages in the Borg and Gall framework. The Alpha Test had 62 participants who were experts, evaluators, and teachers who reviewed the first version of the prototype. In contrast, the Beta Field Trial had 99 participants from five tourism vocational high schools in Bali who assessed the system's functionality in real classroom settings. System usability scale questionnaires, technology acceptance model instruments, and qualitative feedback were used for iterative revisions during both phases. The observations and interviews refined the AI-based e-assessment interface.

TEM Journal. Volume 15, Issue 3, pages 2982-2999, ISSN 2217-8309, DOI: 10.18421/TEM153-79, August 2026.

3.1. Initial Trial (Alpha Test) Results

The purpose of conducting the Alpha Test is to assess the AI-based prototype e-assessment model's basic functions, feasibility, and usability, and to encapsulate the AI-based e-assessment model to assess the Samskara, Cesta, and Krtya levels of competency of tourism vocational high school students in Bali.

The Alpha Test was conducted on 62 respondents, including 2 informatics experts, 2 educational experts, 15 evaluators, and 45 teachers. Several evaluation methods were used, including functional evaluations, heuristic evaluations, think-aloud protocols, and the SUS questionnaires.

The SUS scores of respondents are represented in Table 1. It shows the score on each of the 15 items, and the percentage scores each participant achieved.

Table 1. Alpha test results

Respondent	**Items 1-15**															**Quality Percentage**
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	
Informatics Expert -1	5	4	5	5	4	4	4	5	5	4	4	5	5	5	5	92
Informatics Expert -2	4	5	4	5	4	5	5	5	4	4	5	5	5	5	5	93,33
Educational Expert -1	5	4	4	4	5	5	4	4	5	5	5	4	5	4	4	89,33
Educational Expert -2	4	4	5	4	4	4	5	4	4	5	5	5	4	5	5	89,33
Evaluator-1	5	4	5	5	4	4	5	5	5	5	4	4	5	5	4	92
Evaluator-2	4	4	4	4	5	5	5	4	4	4	5	5	5	4	5	89,33
Evaluator-3	4	4	4	4	4	5	5	5	4	4	5	5	5	5	4	89,33
Evaluator-4	4	4	4	5	4	4	5	5	4	4	5	4	4	4	5	86,67
Evaluator-5	5	4	4	4	5	5	5	4	5	5	5	5	4	4	4	90,67
Evaluator-6	5	4	4	4	4	5	5	4	5	5	5	4	4	5	5	90,67
Evaluator-7	4	4	4	4	5	4	4	4	5	4	4	5	5	5	4	86,67
Evaluator-8	4	4	4	4	5	5	5	4	4	5	4	4	5	5	5	89,33
Evaluator-9	4	4	4	5	5	4	4	4	4	5	5	5	5	4	4	88
Evaluator-10	4	4	4	5	5	5	4	5	5	5	4	4	4	5	5	90,67
Evaluator-11	4	4	4	4	5	4	4	4	5	5	5	4	5	5	5	89,33
Evaluator-12	4	4	4	5	4	4	4	4	4	5	5	4	5	5	5	88
Evaluator-13	4	4	4	5	5	4	4	4	4	5	5	5	4	5	5	89,33
Evaluator-14	4	4	4	5	5	4	4	4	4	4	5	5	4	5	5	88
Evaluator-15	5	4	4	4	5	5	4	4	4	5	4	4	4	5	4	86,67
Teacher -1	5	4	4	4	5	5	4	4	4	5	5	5	4	4	5	89,33
Teacher -2	4	4	4	5	5	4	5	4	4	5	5	4	4	5	4	88
Teacher -3	4	5	4	5	5	4	4	5	5	5	4	5	5	5	5	93,33
Teacher -4	4	5	4	4	5	5	4	4	5	5	4	4	4	5	5	89,33
Teacher -5	4	5	4	4	5	4	4	5	4	4	5	5	5	5	4	89,33
Teacher -6	4	5	4	4	4	5	4	5	5	5	5	4	5	4	4	89,33
Teacher -7	4	5	4	4	4	5	4	5	4	4	4	4	5	5	5	88
Teacher -8	4	5	4	5	5	5	4	4	4	4	4	5	4	4	4	86,67
Teacher -9	4	5	4	4	5	5	4	5	4	5	5	5	4	4	5	90,67
Teacher -10	5	4	4	4	4	5	4	4	4	5	4	4	5	5	5	88
Teacher -11	4	4	5	5	5	5	5	5	4	4	5	4	4	5	5	92
Teacher -12	5	4	5	4	5	5	4	5	4	4	5	5	5	5	4	92
Teacher -13	4	4	4	5	4	4	5	4	5	5	5	4	4	4	5	88
Teacher -14	4	4	4	5	5	5	5	5	4	5	5	5	4	4	4	90,67
Teacher -15	4	4	4	5	4	4	5	4	4	4	5	5	4	5	4	86,67
Teacher -16	5	4	4	5	5	4	5	5	5	5	5	4	5	5	4	93,33
Teacher -17	5	4	4	4	5	5	5	4	4	5	5	4	5	5	5	92
Teacher -18	4	4	4	4	5	4	4	5	5	5	4	4	5	4	4	86,67
Teacher -19	4	4	4	4	4	4	4	4	4	5	5	5	4	4	5	85,33
Teacher -20	5	5	4	4	4	4	5	5	4	5	4	4	5	5	4	89,33

Teacher -21	4	4	5	5	4	5	5	5	4	4	5	5	4	4	4	89,33
Teacher -22	4	5	4	4	4	5	4	4	5	5	5	4	5	4	4	88
Teacher -23	4	5	5	5	4	4	5	4	4	5	5	5	4	4	4	89,33
Teacher -24	5	5	4	4	4	4	5	5	5	5	4	4	4	4	4	88
Teacher -25	5	5	5	4	5	5	5	4	4	4	5	5	4	4	4	90,67
Teacher -26	4	5	5	4	4	5	5	5	4	4	5	5	4	4	5	90,67
Teacher -27	5	5	4	5	4	4	5	5	4	4	5	4	4	5	4	89,33
Teacher -28	5	5	4	4	5	5	5	4	5	5	5	5	4	4	4	92
Teacher -29	4	5	5	4	4	5	5	4	4	4	5	5	4	4	5	89,33
Teacher -30	4	5	4	4	5	4	4	5	5	5	5	4	4	5	4	89,33
Teacher -31	4	4	5	4	5	5	5	5	4	5	4	4	4	4	4	88
Teacher -32	4	4	5	4	5	4	4	4	4	5	5	5	4	5	5	89,33
Teacher -33	5	5	5	4	4	4	4	4	5	4	4	4	5	4	5	88
Teacher -34	4	5	5	4	5	4	5	5	5	4	4	5	4	4	4	89,33
Teacher -35	4	4	5	4	4	4	5	4	4	5	5	5	4	4	4	86,67
Teacher -36	5	5	5	5	5	4	4	5	4	4	5	5	4	4	4	90,67
Teacher -37	4	5	5	4	5	4	4	5	5	5	5	4	4	4	4	89,33
Teacher -38	5	4	4	5	4	5	5	5	4	4	4	5	4	4	4	88
Teacher -39	5	5	5	5	5	4	5	5	5	4	4	4	4	4	4	90,67
Teacher -40	5	4	4	5	4	4	4	5	5	4	5	4	4	4	4	86,67
Teacher -41	5	5	4	5	5	5	5	5	4	5	5	4	4	4	4	92
Teacher -42	4	5	5	5	4	4	5	5	4	5	5	5	4	4	4	90,67
Teacher -43	4	5	4	4	5	5	5	4	4	5	4	4	4	4	4	86,67
Teacher -44	4	4	4	4	4	4	5	5	5	4	4	5	4	4	4	85,33
Teacher -45	4	4	4	5	5	4	5	4	4	5	5	4	4	4	4	86,67
															Mean	**89,23**

As shown in Table 1, the mean SUS score across all 62 respondents was 89.23 percent, which falls within the “Excellent” category. Informatics experts provided the highest ratings with mean scores of 92.00 percent and 93.33 percent, respectively, while educational experts rated the system at 89.33 percent. The 15 evaluators averaged 89.51 percent, and the 45 teachers averaged 89.04 percent. These consistently high scores across all respondent groups indicate strong initial acceptance of the UI design.

Think-aloud protocols and heuristic evaluations identified areas for improvement. The major findings and suggestions for changes are presented in Table 2.

Table 2. Recommendations from Alpha Test

Aspect Tested	Result	Category	Revision Recommendation
Usability (SUS Score)	89,23%	Excellent	Standardize buttons, add icon labels, and add loading indicators
AI Accuracy - Cesta	Motion analysis accuracy	Needs Improvement	Add training data with varied lighting and camera angles; fine-tune model
AI Accuracy - Samskara	Limited expression interpretation	Needs Improvement	Train NLP model with larger dataset containing more local idioms
Scenario Quality - Krtya	Limited options	Needs Improvement	Develop answer options in scenarios; involve teachers to create more realistic scenarios
API Security	Vulnerable	Critical	Improve input validation and API authentication
Teacher Dashboard	Less contextual	Needs Improvement	Add "See Reasoning" (XAI) feature to display basis for AI scoring

The qualitative feedback specifically highlighted that while the overall usability was excellent, several technical and functional aspects required refinement.

The AI accuracy for Cesta (motion analysis) needed improvement through additional training data with varied lighting conditions and camera angles.

Similarly, the NLP model for Samskara required expansion to better interpret local expressions and Balinese cultural context. The Krtya scenarios were found to have limited response options, necessitating collaboration with teachers to develop more realistic tourism-related situations. API security was identified as a critical concern requiring immediate attention.

Teachers also expressed a need for greater transparency in AI decision-making, requesting an Explainable AI feature that would show the reasoning behind scores.

3.2. First Revision Results

Based on the Alpha Test findings and recommendations, systematic revisions were implemented. Figure 1 documents the results of these initial revisions to the AI-based e-assessment model.

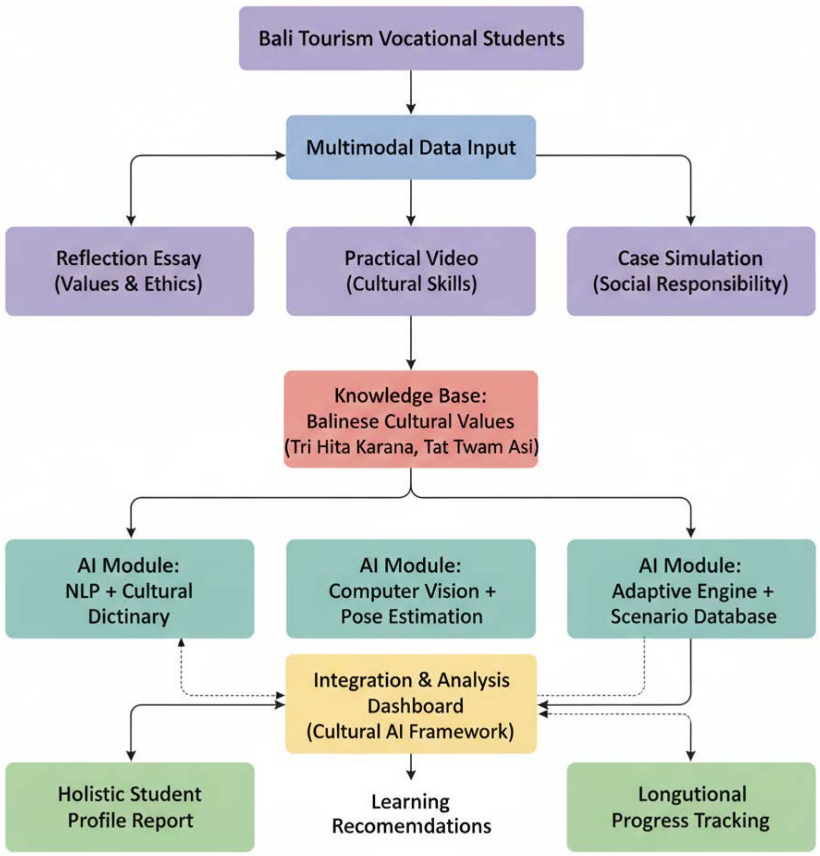

Figure 1. Results of the initial design improvement of the AI-based e-assessment model

The revisions encompassed the following: The standardization of the buttons across the interface, clear labeling of icons where added to enhance icons recognition, loading indicators for system feedback during processing were added; AI model feedback loops were retrained using an expanded dataset; further Krtya scenarios were developed with tourism teachers; API security was strengthened with the addition of input validation and improved authentication; and the explainable AI panel, which outlines the rationale for competency scores was also added.

3.3. Field Trial (Beta Test) Results

The Beta Test took place in real classroom settings at five Tourism SMKs in Bali, involving 99 participants: 2 informatics specialists, 2 education specialists, 30 evaluators, and 65 teachers. This stage assessed user acceptance and system practicality and revealed challenges that were only observed in actual use.

SUS and TAM results from all 99 participants are presented in Table 3.

Table 3. Field trial results

Respondent	Items 1-15															Quality Percentage
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Informatics Expert -1	5	4	4	5	5	4	4	5	4	4	4	5	5	5	5	90,67
Informatics Expert -2	4	5	5	5	4	5	5	5	5	4	5	5	5	5	5	96,00
Educational Expert -1	4	4	5	5	4	4	4	5	5	5	5	4	5	4	4	89,33
Educational Expert -2	4	5	4	4	5	5	5	5	4	5	5	5	4	5	5	93,33
Evaluator-1	4	5	5	5	5	4	5	4	4	5	5	4	4	5	4	90,67
Evaluator-2	4	5	4	4	4	4	4	5	5	5	4	5	5	5	5	90,67
Evaluator-3	4	4	4	4	4	5	4	4	5	5	4	4	4	5	5	86,67
Evaluator-4	4	5	4	5	5	5	4	5	4	4	5	5	5	5	4	92,00
Evaluator-5	4	4	5	5	5	4	5	5	5	5	5	4	5	4	4	92,00
Evaluator-6	4	4	4	5	5	4	5	5	5	4	4	4	5	5	5	90,67
Evaluator-7	4	4	4	5	5	5	4	5	5	4	4	5	4	4	4	88,00
Evaluator-8	4	4	4	4	5	5	4	5	5	5	5	5	4	4	5	90,67
Evaluator-9	4	4	4	5	4	4	4	4	5	5	5	4	5	5	5	89,33
Evaluator-10	4	4	4	5	5	5	4	4	4	5	5	4	5	5	5	90,67
Evaluator-11	5	4	4	5	5	4	4	4	4	5	5	5	4	5	5	90,67
Evaluator-12	4	5	5	5	4	5	4	4	4	4	5	5	4	5	5	90,67
Evaluator-13	5	5	5	4	4	5	4	4	4	5	4	4	4	5	4	88,00
Evaluator-14	4	4	5	5	5	4	4	5	4	5	5	5	4	4	5	90,67
Evaluator-15	5	4	5	4	4	5	5	4	4	5	5	4	4	5	4	89,33
Evaluator-16	5	4	4	5	5	4	4	4	5	5	4	5	5	5	5	92,00
Evaluator-17	4	5	5	5	4	5	4	4	4	4	5	5	5	4	5	90,67
Evaluator-18	4	4	5	5	5	4	4	4	4	4	5	5	5	5	4	89,33
Evaluator-19	5	5	5	4	4	4	4	4	4	4	5	4	4	4	5	86,67
Evaluator-20	4	4	4	5	5	4	4	4	5	5	5	5	4	4	4	88,00
Evaluator-21	5	4	4	4	4	5	5	4	5	5	5	4	4	5	5	90,67
Evaluator-22	4	4	4	4	5	4	4	4	5	4	4	5	5	5	4	86,67
Evaluator-23	4	4	4	4	5	5	5	4	4	5	4	4	5	5	5	89,33
Evaluator-24	4	4	4	5	5	4	4	4	4	5	5	5	5	4	4	88,00
Evaluator-25	4	4	4	5	5	5	4	5	5	5	4	4	4	5	5	90,67
Evaluator-26	4	4	4	4	5	4	4	4	5	5	5	4	5	5	5	89,33
Evaluator-27	4	4	4	5	4	4	4	4	4	5	5	4	5	5	5	88,00
Evaluator-28	4	4	4	5	5	4	4	4	4	5	5	5	4	5	5	89,33
Evaluator-29	4	4	4	5	5	4	4	4	4	4	5	5	4	5	5	88,00
Evaluator-30	5	4	4	4	5	5	4	4	4	5	4	4	4	5	4	86,67
Teacher -1	5	4	4	4	5	5	4	4	4	5	5	5	4	4	5	89,33
Teacher -2	4	4	4	5	5	4	5	4	4	5	5	4	4	5	4	88,00
Teacher -3	4	5	4	5	5	4	4	5	5	5	4	5	5	5	5	93,33
Teacher -4	4	5	4	4	5	5	4	4	5	5	4	4	4	5	5	89,33
Teacher -5	4	5	4	4	5	4	4	5	4	4	5	5	5	5	4	89,33
Teacher -6	4	5	4	4	4	5	4	5	5	5	5	4	5	4	4	89,33
Teacher -7	4	5	4	4	4	5	4	5	4	4	4	4	5	5	5	88,00
Teacher -8	4	5	4	5	5	5	4	4	4	4	4	5	4	4	4	86,67
Teacher -9	4	5	4	4	5	5	4	5	4	5	5	5	4	4	5	90,67
Teacher -10	5	4	4	4	4	5	4	4	4	5	4	4	5	5	5	88,00
Teacher -11	4	4	5	5	5	5	5	5	4	4	5	4	4	5	5	92,00
Teacher -12	5	4	5	4	5	5	4	5	4	4	5	5	5	5	4	92,00
Teacher -13	4	4	4	5	4	4	5	4	5	5	5	4	4	4	5	88,00
Teacher -14	4	4	4	5	5	5	5	5	4	5	5	5	4	4	4	90,67
Teacher -15	4	4	4	5	4	4	5	4	4	4	5	5	4	5	4	86,67
Teacher -16	5	4	4	5	5	4	5	5	5	5	5	4	5	5	4	93,33
Teacher -17	5	4	4	4	5	5	5	4	4	5	5	4	5	5	5	92,00
Teacher -18	4	4	4	4	5	4	4	5	5	5	4	4	5	4	4	86,67
Teacher -19	4	4	4	4	4	4	4	4	4	5	5	5	4	4	5	85,33
Teacher -20	5	5	4	4	4	4	5	5	4	5	4	4	5	5	4	89,33
Teacher -21	4	4	5	5	4	5	5	5	4	4	5	5	4	4	4	89,33
Teacher -22	4	5	4	4	4	5	4	4	5	5	5	4	5	4	4	88,00
Teacher -23	4	5	5	5	4	4	5	4	4	5	5	5	4	4	4	89,33
Teacher -24	5	5	4	4	4	4	5	5	5	5	4	4	4	4	4	88,00
Teacher -25	5	5	5	4	5	5	5	4	4	4	5	5	4	4	4	90,67

Respondent	Items 1-15															Quality Percentage
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Teacher -26	5	5	5	4	4	5	4	4	5	5	5	4	4	5	4	90,67
Teacher -27	4	4	5	5	5	4	4	5	4	4	5	5	5	4	4	89,33
Teacher -28	5	4	5	4	4	5	5	4	5	4	5	4	4	5	5	90,67
Teacher -29	5	4	4	5	5	4	4	4	5	4	4	5	5	4	4	88,00
Teacher -30	4	5	5	5	4	5	4	4	5	5	5	4	4	5	4	90,67
Teacher -31	4	4	5	5	5	4	4	4	4	4	5	5	5	4	4	88,00
Teacher -32	5	5	5	4	4	4	4	4	5	4	5	4	4	5	5	89,33
Teacher -33	4	4	4	5	5	4	4	4	5	4	4	5	5	4	4	86,67
Teacher -34	4	4	5	4	5	5	5	5	5	5	4	4	5	4	4	90,67
Teacher -35	4	4	5	4	5	4	4	4	4	5	5	5	4	4	5	88,00
Teacher -36	5	5	5	4	4	4	4	5	4	5	4	4	5	5	4	89,33
Teacher -37	4	5	5	4	5	4	5	5	4	4	5	5	4	4	4	89,33
Teacher -38	4	4	5	4	4	4	5	4	5	5	5	4	5	4	4	88,00
Teacher -39	5	5	5	4	4	5	4	4	4	5	5	5	4	4	4	89,33
Teacher -40	4	4	5	5	5	4	4	5	5	5	4	4	4	4	4	88,00
Teacher -41	5	4	5	4	4	5	5	4	4	4	5	5	4	4	4	88,00
Teacher -42	5	4	4	5	5	4	4	4	5	4	5	4	4	5	4	88,00
Teacher -43	4	5	4	4	5	4	5	5	4	5	4	4	4	4	4	86,67
Teacher -44	5	4	4	4	4	4	4	4	5	4	4	4	4	5	4	84,00
Teacher -45	4	4	4	4	5	4	5	5	4	5	4	4	4	5	4	86,67
Teacher -46	5	4	4	4	5	4	5	4	5	5	5	5	5	4	4	90,67
Teacher -47	5	5	5	5	4	4	5	5	4	5	4	5	4	4	4	90,67
Teacher -48	5	4	5	4	4	4	4	4	4	4	4	5	5	5	4	86,67
Teacher -49	4	4	5	5	5	4	5	5	4	4	4	5	4	4	5	89,33
Teacher -50	4	5	4	4	5	4	4	5	5	5	5	4	4	5	4	89,33
Teacher -51	4	4	5	4	5	5	5	5	4	5	4	4	4	4	4	88,00
Teacher -52	4	4	5	4	5	4	4	4	4	5	5	5	4	5	5	89,33
Teacher -53	5	5	5	4	4	4	4	4	5	4	4	4	5	4	5	88,00
Teacher -54	4	5	5	4	5	4	5	5	5	4	4	5	4	4	4	89,33
Teacher -55	4	4	5	4	4	4	5	4	4	5	5	5	4	4	4	86,67
Teacher -56	5	5	5	5	5	4	4	5	4	4	5	5	4	4	4	90,67
Teacher -57	4	5	5	4	5	4	4	5	5	5	5	4	4	4	4	89,33
Teacher -58	5	4	4	5	4	5	5	5	4	4	4	5	4	4	4	88,00
Teacher -59	5	5	5	5	5	4	5	5	5	4	4	4	4	4	4	90,67
Teacher -60	5	4	4	5	4	4	4	5	5	4	5	4	4	4	4	86,67
Teacher -61	5	5	4	5	5	5	5	5	4	5	5	4	4	4	4	92,00
Teacher -62	4	5	5	5	4	4	5	5	4	5	5	5	4	4	4	90,67
Teacher -63	4	5	4	4	5	5	5	4	4	5	4	4	4	4	4	86,67
Teacher -64	4	4	4	4	4	4	5	5	5	4	4	5	4	4	4	85,33
Teacher -65	4	4	4	5	5	4	5	4	4	5	5	4	4	4	4	86,67
Mean																89,20

The average SUS score for all 99 participants in the study was 89.20%, putting it in the “Excellent” category. Informatics specialists did better with 90.67 percent and 96.00 percent. Educational specialists scored the system at 89.33 and 93.33. Of the 30 evaluators, 89.42 percent scored, while of the 65 teachers, 88.95 percent.

According to the technology acceptance model, the average perceived usefulness and perceived ease of use scores were 89.20%, indicating high acceptance in both cases.

Real-world implementation challenges and user recommendations were captured through structured observations and in-depth interviews. Table 4 captures the suggested revisions and the findings from the field trial.

Table 4. Summary of field trial results and revision recommendations

Aspect Tested	Result	Category	Revision Recommendation
Usefulness	89.20%	Excellent	Maintain
Ease of Use	89.20%	Excellent	Optimize for slow internet & low specifications; add video tutorials
Usability (SUS Score)	89.20%	Excellent	Simplify XAI explanations; improve UI responsiveness
AI Accuracy - Cesta	Good, but needs expansion	Improvement	Add video pre-processing; provide video recording guidelines
AI Accuracy - Samskara	Good, but needs expansion	Improvement	Train model with informal text; add spell check
Scenario Quality - Krtya	Good, but needs expansion	Improvement	Expand scenario bank and answer options
Feedback Feature	Less actionable	Needs Improvement	Add AI-generated feedback suggestions
Reporting Mechanism	None	Critical	Add "Report Inaccuracy" button

The field trial demonstrated that high usability ratings for the system were insufficient to overcome the challenges encountered in actual classroom use. Teachers reported that during use, the system slowed and lagged due to slow Internet connections or device specifications. The XAI explanations, although appreciated, were often too complex for non-technical users. The AI analysis for Cesta was good, but it needed to be expanded to cover different video recording conditions. Teachers requested guidance for video recording to meet the analysis requirements. The Samskara NLP model performed well; however, informal student language, often containing spelling errors, was a problem. There could be an integrated spell-checker, which might also alleviate the problem of inadequate training in student writing patterns.

Teachers wanted more Krtya scenarios with branching options to simulate real tourism situations better. The feedback was valuable, but the criticism was unfounded. Teachers wanted AI feedback to drive changes, rather than just a verbal description of what could be changed. The most critical problem was the lack of a mechanism for users to report what they believed were errors in the AI evaluation. It was, and still is, often seen as an essential element for building user confidence in the system.

3.4. Final Revision Results

The comprehensive field trials have enabled one final round of adjustments, completing the first iteration of the production-ready prototype. The last set of changes is shown in the final revised design in Figure 2.

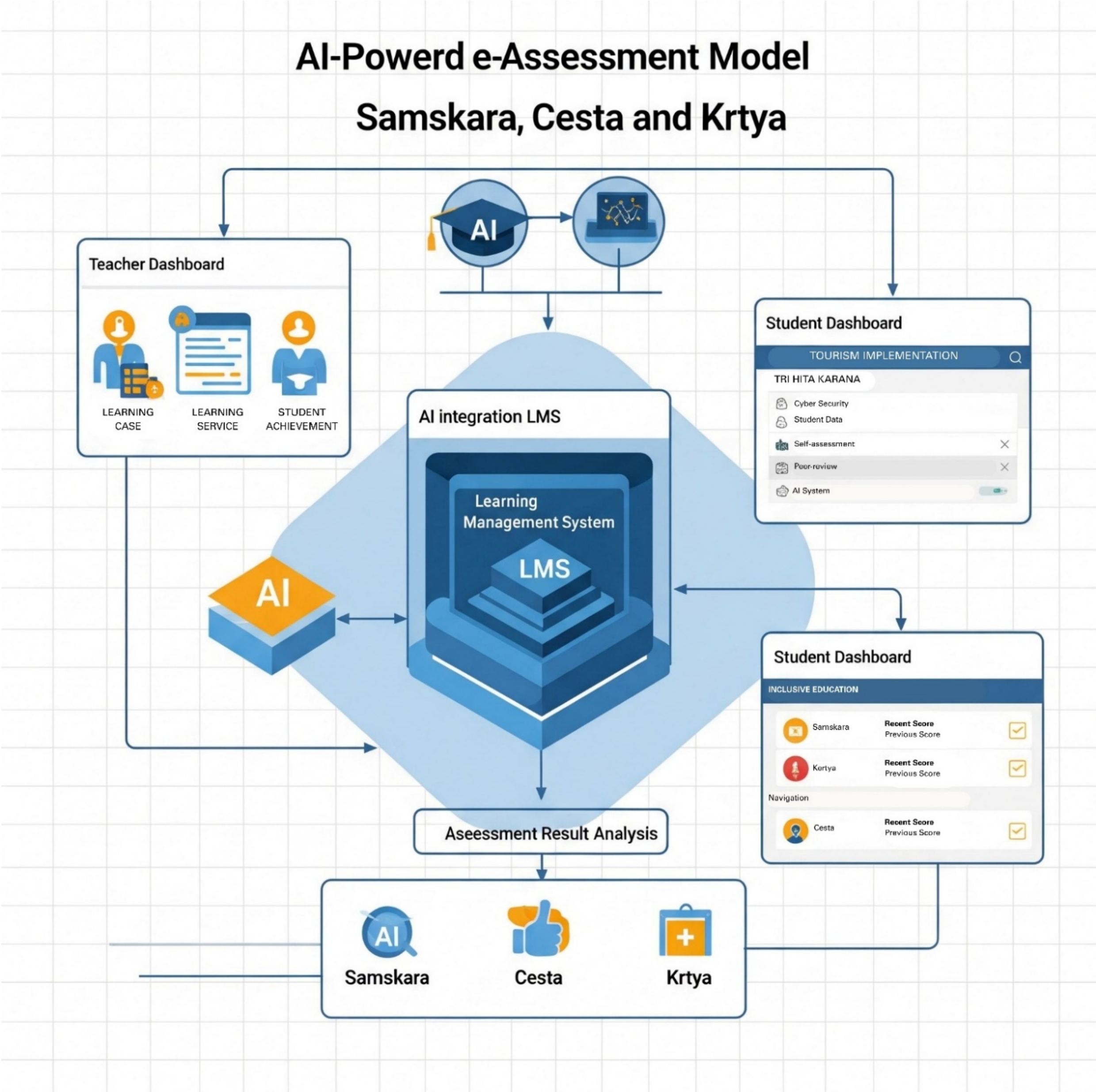

Figure 2. Design improvement results after field trial of AI-based e-assessment

Lazy loading, video compression, and an offline mode were also introduced as additional optimizations for low-bandwidth environments. The ease of use for AI Explanations (XAI) was simplified by layering information so a user can choose to view a more basic rationale or an additional, more detailed explanation. A new real-time video recording guide was added to demonstrate optimal lighting, angles, and distance. Other enhancements included improved spell-checking and further training of the NLP writing model using student writing samples. The Krtya scenario bank was expanded to over 50 scenarios with multiple branching options. There are now AI suggestions for how users can improve their scenarios. Also, users can now directly Report Inaccuracy issues to the system's database to improve system updates.

4. Discussion

The results from the Alpha and Beta Tests show that the proposed AI-based e-assessment UI achieved very high scores on usability and user acceptance. Alpha Test results showed a mean SUS score of 89.23 percent, and Field Trial results showed a 89.20 percent SUS score, placing them both in the “Excellent” category according to the guidelines from [28]. These results are significantly above the average acceptability score of 70 percent and fall on the positive end of the range reported in top-quartile educational dashboards in the literature [29]. Measurements from the technology acceptance model also support this finding, as both perceived usefulness and perceived ease of use scored 89.20 percent in the field trial.

These results stand out because they demonstrate usability while incorporating difficult cultural aspects and sophisticated AI components. Many assume that, as a result of cultural embellishment, functional clarity will be sacrificed. The results, however, reinforce the cultural UX theory, which states that when design and functionality align with users' socio-cultural backgrounds, cognitive load is reduced and emotional involvement increases [30]. Balinese teachers and evaluators in the study described feeling "at home" with the interface. Quite a few mentioned that the technology became less unfamiliarand more a part of their daily cultural experience because of the traditional Balinese-like motifs and earthy colour palette. Their emotional connection explains the high acceptance scores. Users evaluated the interface as a system that valued and enhanced their cultural and civilizational legacy rather than merely a functional tool. Consistency in scores across specific respondent categories, such as informatics specialists, education specialists, evaluators, and instructors, is commendable. Although subject matter experts are expected to provide more pessimistic evaluations, their scores were surprisingly high, suggesting that the system fulfills its functionality, as well as its teaching and socio-cultural concerns. Given the slight score dispersion from 85.33 percent to 96.00 percent, the design succeeded in achieving numerous stakeholder perspectives, an essential factor for the system to be employed across a wide range of educational contexts.

The study's successful outcome identifies deep cultural integration as a substantive, rather than ornamental, design principle. The decision to weave Balinese Tri Hita Karana philosophy into all facets of the interface, from the design of the login page to the visualizations on the dashboard and the competency frameworks, was certainly not for mere decoration. It was a constitutive factor in determining the nature of user-system interactions and the level of user trust in the system. Over the years, the cultural UI design literature has highlighted the roles of language, structure, and icons in designing culturally inclusive interfaces [23]. However, this literature has mostly focused on the translation of language and visual design elements, without addressing deeper levels of integration. This study illustrates that, from the perspectives of users and teachers, integrating certain indigenous philosophical ideas (Samskara, Cesta, Krtya) into the design and, in this case, the assessment and construction, leads users to feel a strong sense of ownership and positivity. Teachers explained their pride in how their indigenous wisdom has been measured and transformed into a contemporary digital expression.

Contrary to the perception that adopting technology undermines cultural preservation, this study demonstrates that intentionally and collaboratively developed technologies can revitalize culture. The system's dashboard features are designed not to impose Western values in place of local traditional values; rather, they express local traditional values in a way that is accessible and understandable to modern digital natives. In this light, the system transforms the euphoria of the risks of modernization into an opportunity, allowing the competencies of Samskara, Cesta, and Krtya to remain alive and measurable, and not to lose themselves in the realm of the purely philosophical. The efficacy of such cultural fusion goes far beyond Bali. It provides a model for indigenous and minority populations around the world that are still being negatively homogenized by globalized technologies. It shows that by embedding cultural experts in the design, restructuring the use of local philosophical systems into operationalizable indicators, and incorporating local design into the visual and behavioral interface, it is possible to develop an artificial system that neither ignores local wisdom nor erases it, but rather uplifts it.

Despite excelling in the system's cultural and usability aspects, the AI's accuracy raised significant issues that required multiple rounds of improvement. The Alpha Test showed that Cesta's motion-tracking analysis struggled to perform adequately in certain lighting conditions and that Samskara's NLP had issues with localized idioms. These limitations highlight an important reality about AI in the educational evaluation process: the effectiveness of an algorithm is determined by the quality of the information used to train it, which reflects how it is used in the world [31]. The first model used to create the Cesta analysis function was, in all likelihood, trained on video data to establish a baseline of steady lighting and unobstructed camera viewpoints. This scenario completely ignores the realities of Balinese classrooms, where lighting and recording devices (audio and video) may not be placed in the same area or configured at the same camera position as in other classrooms. Students may not record with a camera. Moreover, the Samskara NLP encountered issues with informal and non-standard student language, including Balinese slang and other colloquialisms, misspellings, and variations in non-standard language. These issues are profoundly cultural, contextual, and technical. Following the Alpha Test, revisions included retraining models across various lighting conditions and angles, building datasets with regional vernacular, and involving educators in scenario building. It exemplifies the participatory AI design approach, recognized as critical for equitable AI design [32].

By positioning educators as more than users but as co-designers and integrating their contextual expertise, the system's ecological validity increased tremendously. The trials indicated "Good" AI accuracy, with expansion needed, suggesting considerable advancements but requiring further iterative work. There are critical lessons for AI-based competence assessment. The assessment of non-cognitive skills, whether in engineering design, social-emotional learning, or other relevant frameworks, is grounded in context. An AI model trained on Western datasets will fail to identify a lack of empathy, unethical reasoning, or poor collaborative practices among Balinese students, as those behaviours are culturally contextual. This research highlights the pressing need for culturally specific AI frameworks, in contrast to the culturally myopic view of Western data sets, and the expectation of equal accuracy across all contexts [33].

By far, the most consistent finding in both phases of the study is the need for AI explanation in operational decision-making. As a result of the Alpha Test, the teaching staff described the dashboard as 'less contextual' and have requested the ability to 'see the reason' for AI scores, prompting the addition of an explainable AI (XAI) feature to the dashboard. During the field trial, although the XAI feature was well received, there was a request for a simpler explanation, indicating that the first iteration of XAI was technically correct but too complex for casual users. The aforementioned comments provide additional support for the emerging body of literature on algorithm aversion and the need for AI trust [34]. For users in vulnerable situations, such as educational assessment, AI-made judgments are not accepted. Thus, there is a lack of the fundamental trust required for technology adoption. The primary challenge for developers will be to maintain transparency (trust) without compromising functionality. What is clear to a researcher may be completely obscure to a classroom teacher. The final revisions that reduce XAI user accessibility while retaining its layered functionality indicate the increasing value placed on balancing theoretical pragmatism and cohesiveness. The trust-building function of explainable artificial intelligence (XAI) would be more critical for culturally sensitive assessments. With AI grading assessments that involve constructive feedback in critical, thoughtful, and reflective (Samskara and Cesta) dimensions, teachers need to access the system to understand the evaluation reasoning, not just the score. The XAI panel, displaying keyword triggers, behavior, and evaluation patterns across benchmarks, enabled teachers to determine whether the AI was actually measuring the dimensions it claimed to measure, as well as to align assessments with the norms of the given culture.

In interviews, several teachers stated that the reasoning the AI extended helped them articulate the critical, thoughtful, and reflective (Samskara) dimensions of their teaching practice descriptions, indicating that XAI has developmental functions that transcend transparency.

The field trial highlighted several issues stemming from operational circumstances that occur exclusively in genuine educational settings. It underscores the need to conduct beta testing rather than testing in a controlled environment during an alpha phase. Significant issues revealed were the system's inconsistent performance on low-bandwidth, low-spec devices and its inability to capture user feedback, allow users to provide feedback, and report system inadequacies. The described issues stem from the gaps in digital technology. In Indonesia, even in areas deemed semi-urban or developed, there are blatant gaps of this kind that the country is currently grappling with. Where schools in Denpasar have reliable internet and modern computing devices, schools in the hinterlands of Bali are enduring poor internet connectivity and obsolete computing devices. The system's performance, meaning its adaptability to high bandwidth, is aimed at increasing the gaps between schools and addressing technology inequity. The implemented changes, lazy loading, video compression, and offline capabilities, represent a strong inclination to adopt design standards that prioritize accessibility over design standards that prioritize optimum performance. It supports the demand for "appropriate technology" in economically disadvantaged or developing areas, where systems must be adapted to local infrastructure. It contrasts with the idea of "importing" technology from developed countries without considering local conditions and needs [36]. The addition of a "Report Inaccuracy" button shows that broken trust can be repaired and underscores the need for the feature. Users of AI systems, particularly in education, need to feel they are not just passive recipients of AI judgment but part of an iterative process. There are several incentives for designing a report button. For example, it empowers teachers to provide system feedback. It also provides developers with data for model updates, shows that developers understand the AI is wrong, and invites people to get involved. Field trial participants felt the absence of mechanisms to collect user feedback during the AI system design process was 'Critical', supporting the notion that user agency is not optional in AI-supported testing; it is a requirement.

Euforia Modernization refers to an extreme enthusiasm for Modernization and often displays a negative view of the uncritical, rampant adoption of foreign technologies, which may result in the loss of a culturally and morally unified society [13]. Even so, the research conducted through this study has recorded this phenomenon in a more positive context.

In this case, positive context, Modernization is a positive force for culturally conscious and collaborative culturally conscious and collaborative planning of Modernization to preserve. The teachers and students who interacted with the system were not alienated from it, but it did not alienate the system from their cultural system. The system reflected a creative reprisal of contextualized local wisdom adapted to modernization. It prioritized the community's own regulatory values, ensuring that Western metrics did not overshadow traditional ways of knowing. Rather than cultural erosion, this is cultural evolution. This change has implications for the way people think about the intersection of tech and age-old customs. The distinction between "new" and "old," which is often the focus of discussions of modernity, may be misleading. This research shows the potential of the third option, which is modern in structure and old in essence. The tech involves advanced AI, NLP, machine learning, and interactive dashboards, but the values and the design sheaves are anchored in Balinese culture. This combination, rather than unqualified tradition or unqualified modernity, may prove the most enduring route for societies facing the challenge of rapidly changing technologies.

5. Conclusion

The purpose of this study was to create and confirm the cultural validity of an AI-based e-assessment user interface that assesses the three competencies of Samskara, Cesta, and Krtya of Bali's tourism vocational high school students. The final iteration of the interface, after multiple rounds of Alpha and Beta testing, received very positive scores for usability and acceptance from 161 participants across five schools. The testing results support the idea that robust cultural integration of an interface does not inhibit its technological capabilities. The revisions to the AI accuracy, scenario richness, XAI transparency, low-bandwidth usage, and user feedback instruments yielded a production-ready interface that showcased a high level of AI while giving great consideration to the Tri Hita Karana of Balinese philosophy. The study also demonstrated that the operationalization of AI-based measurement of non-cognitive indigenous competencies/values was feasible. An important finding of the study was that an explainable AI mechanism was imperative for trust and acceptance of the solution by the teachers. The study also examined the culturally appropriate use of technology, suggesting that when technology is purposefully designed with local values in mind, it can be seen as a means to safeguard culture rather than erode it.

Although this research is specifically focused on the Balinese vocational context and requires longitudinal validation of student outcomes, it offers a replicable, foundational model for culturally attuned AI design, encouraging indigenous communities globally to integrate technology while retaining their foundational values. In adapting this model to other cultural contexts and educational systems, future research is needed on the sustainability of student competency growth and on student and assessment experience.

Acknowledgements

The author wishes to acknowledge with gratitude the Directorate General of Research and Development, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, for the funding received in this research project. This grant, with Contract Number 2166/LL8/AL.04/2025, dated June 5, 2025, goes a long way in facilitating the publication of this article.

References:

[1]. Sibley, L., et al. (2025). Adaptive teaching with technology enhances lasting learning. *Learning and instruction*, *99*, 102141. Doi: 10.1016/j.learninstruc.2025.102141.

[2]. Jin, T. (2025). Methods and reliability study of moral education assessment in universities: A machine learning-based approach. *Alexandria Engineering Journal*, *125*, 20-28. Doi: 10.1016/j.aej.2025.03.095.

[3]. Playán, E., et al. (2024). Irrigation modernization in the Ebro – Aragón region of Spain: Past and future trends. *Agricultural Water Management*, *302*, 108975. Doi: 10.1016/j.agwat.2024.108975.

[4]. Dubicki, E. (2019). Mapping curriculum learning outcomes to ACRL's Framework threshold concepts: A syllabus study. *The Journal of Academic Librarianship*, *45*(3), 288-298. Doi: 10.1016/j.acalib.2019.04.003.

[5]. Gal, C., & Ryder, C. H. (2025). Unlocking potential: Comparing collaborative and traditional learning methods for students with learning disabilities in special education classrooms. *Social Sciences & Humanities Open*, *11*, 101521. Doi: 10.1016/j.ssaho.2025.101521.

[6]. Hidalgo, A. (2024). Your room is ready: Tourism and urban revival. *Regional Science and Urban Economics*, *109*, 104059. Doi: 10.1016/j.regsciurbeco.2024.104059.

[7]. Divayana, D. G. H., et al. (2021). DIVAYANA formula for determining priority ranking the recommendations in evaluation applications. *IOP Conference Series: Materials Science and Engineering*, 1098(3), 032063. Doi: 10.1088/1757-899X/1098/3/032063.

[8]. Suyasa, P. W. A., et al. (2025). Field trial of Provus-Alkin-amalgamation evaluation application based on Weighted-Product-Rwa-Bhineda mods. *Journal of Education and Learning (EduLearn)*, *19*(1), 495-505. Doi: 10.11591/edulearn.v19i1.21114.

[9]. Fischer, S., & Barabasch, A. (2023). Conceptualizations and implementation of creativity in higher vocational teacher education–a qualitative study of lecturers. *Empirical Research in Vocational Education and Training*, *15*(1), 6. Doi: 10.1186/s40461-023-00144-y.

[10]. Fereidani, B. M., & Üçtuğ, F. G. (2024). Towards sustainable production of sesame products: Comparison of traditional and modern production systems via a life cycle assessment approach. *Cleaner and Responsible Consumption*, *12*, 100166. Doi: 10.1016/j.clrc.2023.100166.

[11]. Yang, F. (2025). Computer interactive intelligent music therapy assisted student mental health information processing in life sciences. *SLAS technology*, *32*, 100269. Doi: 10.1016/j.slast.2025.100269.

[12]. Ng, Z. J., et al. (2025). Development of the student emotion regulation assessment (SERA) for children and adolescents in grades 1–12. *Social and Emotional Learning: Research, Practice, and Policy*, *5*, 100104. Doi: 10.1016/j.sel.2025.100104.

[13]. Liu, C. Y. (2025). Assessment framework for students' STEAM performance: Teachers' interviews and classroom observations in Taiwan. *Teaching and Teacher Education*, *165*, 105135. Doi: 10.1016/j.tate.2025.105135.

[14]. Magruder, M. L., et al. (2025). Comparison of ChatGPT plus (version 4.0) and pretrained AI model (Orthopod) on orthopaedic in-training exam (OITE). *The Surgeon*, *23*(3), 187-191. Doi: 10.1016/j.surge.2025.04.004.

[15]. Lachheb, A., Leung, J., Abramenka-Lachheb, V., & Sankaranarayanan, R. (2025). AI in higher education: A bibliometric analysis, synthesis, and a critique of research. *The Internet and Higher Education*, *67*, 101021. Doi: 10.1016/j.iheduc.2025.101021.

[16]. Fošner, A., & Aver, B. (2025). AI chatbots in higher education: Students' beliefs and concerns. *Sustainable Futures*, *9*, 100734. Doi: 10.1016/j.sftr.2025.100734.

[17]. Pikhart, M., & Al-Obaydi, L. H. (2025). Reporting the potential risk of using AI in higher Education: Subjective perspectives of educators. *Computers in Human Behavior Reports*, *18*, 100693. Doi: 10.1016/j.chbr.2025.100693.

[18]. Kumar, G., Sharma, D., & Bhardwaj, B. (2025). The future of work and education in AI-driven innovative systems: A systematic literature review and lexicometric analysis. *The International Journal of Management Education*, *23*(3), 101221. Doi: 10.1016/j.ijme.2025.101221.

[19]. Villaceran, E. C., & Himang, C. M. (2025). Data on behavioural intention to use AI copilot through TAM and AI ecological education policy lens. *Data in Brief*, *61*, 111686. Doi: 10.1016/j.dib.2025.111686.

[20]. Liu, C. C., Wang, H. J., & Gu, X. Q. (2025). From access to mastery: Integrating AI in blended learning for equitable, inclusive, and accessible music theory educations. *The Internet and Higher Education*, *66*, 101018. Doi: 10.1016/j.iheduc.2025.101018.

[21]. Kajiwara, Y., Matsuoka, A., & Shinbo, F. (2023). Machine learning role playing game: Instructional design of AI education for age-appropriate in K-12 and beyond. *Computers and Education: Artificial Intelligence*, *5*, 100162. Doi: 10.1016/j.caeai.2023.100162.

[22]. Fan, Y. L., et al. (2025). Enhancing safety with an AI-empowered assessment and monitoring system for BSL-3 facilities. *Heliyon*, *11*(1). Doi: 10.1016/j.heliyon.2024.e40855.

[23]. Khaliq, Z., Khan, D. A., & Farooq, S. U. (2023). Using deep learning for selenium web UI functional tests: A case-study with e-commerce applications. *Engineering Applications of Artificial Intelligence*, *117*, 105446. Doi: 10.1016/j.engappai.2022.105446.

[24]. Fadlillah, M., Rahman, B., & Istiq'faroh, N. (2024). Development of teaching materials with agricultural insight to introduce of agriculture in primary school students. *Journal of Education and Learning (EduLearn)*, *18*(4), 1224-1233. Doi: 10.11591/edulearn.v18i4.21641.

[25]. Agung, D. A. G., Nasih, A. M., & Kurniawan, B. (2024). Local wisdom as a model of interfaith communication in creating religious harmony in Indonesia. *Social Sciences & Humanities Open*, *9*, 100827. Doi: 10.1016/j.ssaho.2024.100827.

[26]. Akakpo, M. G., Ahardy, D. O., & Kumankumah, S. S. (2025). Towards digital information literacy guidelines for African libraries: a survey of the relationship between digital and information literacy of university students in Ghana. *The Journal of Academic Librarianship*, *51*(4), 103067. Doi: 10.1016/j.acalib.2025.103067.

[27]. Saikali, S., et al. (2025). Development and assessment of an AI-based machine learning model for predicting urinary continence and erectile function recovery after robotic-assisted radical prostatectomy: insights from a prostate cancer referral center. *Computer methods and programs in biomedicine*, *259*, 108522. Doi: 10.1016/j.cmpb.2024.108522.

[28]. Ardana, I., Ariawan, I., & Sugiharni, G. A. D. (2021). The Expansion of Sociocultural Theory-Oriented Mathematical Learning Model. *Cypriot Journal of Educational Sciences*, *16*(6), 3016-3032. Doi: 10.18844/cjes.v16i6.6493.

[29]. Low, M. P., et al. (2025). Beyond conventional adoption: new insights into AI facilitators in Higher Education Institutions. *Sustainable Futures*, *9*, 100781. Doi: 10.1016/j.sftr.2025.100781.

[30]. Wiese, L. J., et al. (2025). AI ethics education: A systematic literature review. *Computers and Education: Artificial Intelligence*, *8*, 100405. Doi: 10.1016/j.caeai.2025.100405.

[31]. Guan, L., Lee, J. C. K., Zhang, Y., & Gu, M. M. (2025). Investigating the tripartite interaction among teachers, students, and generative AI in EFL education: A mixed-methods study. *Computers and Education: Artificial Intelligence*, *8*, 100384. Doi: 10.1016/j.caeai.2025.100384.

[32]. Wiss, A., et al. (2025). Development of the AI Acceptance Scale for Interprofessional Education (AAIPE) and collaborative practice settings. *Journal of Interprofessional Education & Practice*, *40*, 100752. Doi: 10.1016/j.xjep.2025.100752

[33]. Entorf, A. K., & Dohmen, T. J. (2025). The effect of compulsory education on non-cognitive skills: Evidence from low-and middle-income countries. *Economics of Education Review*, *107*, 102654. Doi: 10.1016/j.econedurev.2025.102654.

[34]. Obuobi, B., et al. (2025). Comparing energy conservation behavior from the perspective of the karma effect. *Renewable and Sustainable Energy Reviews*, *217*, 115753. Doi: 10.1016/j.rser.2025.115753.

[35]. Suryawan, I. W. K., et al. (2025). Evaluating tourist preferences influences on renewable energy initiatives in coastal tourism settings. *Regional Studies in Marine Science*, *85*, 104148. Doi: 10.1016/j.rsma.2025.104148.

[36]. Syamala, S., Shrivastava, A., & Karmakar, S. (2025). Identifying interface design factors impacting user experience in digital learning platforms-A pilot study. *Social Sciences & Humanities Open*, *11*, 101662. Doi: 10.1016/j.ssaho.2025.101662.