JURNAL EDUSCIENCE (JES)
p-ISSN : 2303 - 355X | e-ISSN : 2685 - 2217
PUBLISHED BY : LPPM of UNIVERSITAS LABUHANBATU

AI-Integrated DEEPIR Model-Based Student Worksheet: Enhancing Learning Outcomes and HOTS in Higher Education

Budi Halomoan Siregar[1], Zul Amry[2], Kairuddin[3], Cut Latifah Zahari[4], Noorhelyna Razali[5]

[1, 2, 3]Departement of Mathematics Education, Universitas Negeri Medan, Sumatra Utara, Indonesia
[4]Department of Mathematics Education, Universitas Muslim Nusantara Al-Washliyah, Sumatra Utara, Indonesia
[5]Department of Engineering Education, 43600 Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
*budihalomoan@unimed.ac.id

ARTICLE INFO

Keywords:
Artificial Intelligence
DEEPIR Model
Outcome
HOTS

ABSTRACT

Purpose – This study aims to develop student worksheets (SW) by integrating Artificial Intelligence (AI) into the DEEPIR model, aligning with an outcome-oriented instructional framework to improve HOTS and learning products (innovative learning media and scientific publications).

Methodology – This study employed a design research approach, implemented through three main stages: preparation and design, design experiment, and retrospective analysis. The research participants were 22 students of Mathematics Education enrolled in the Learning Media course. The research instruments included observation sheets, validation and practicality sheets, HOTS-based tests, and product assessment rubrics. Data were obtained through learning observations and analysis of the validity, practicality, effectiveness, and results of student products. All data were analyzed descriptively and inferentially to assess the quality of the student worksheets.

Findings – The study's results demonstrate high levels of validity, practicality, and effectiveness. The n-gain value is 0.783 (high category). All students (100%) successfully produced innovative learning media products that met the criteria for content and pedagogical validity. In addition, 82% of students published scientific articles in nationally accredited journals with a minimum Sinta 4 rating. These findings indicate that AI-integrated DEEPIR-based student worksheets significantly improve HOTS and students' ability to produce media products and scientific articles.

Contribution – This study produce a replicable model that educators can adapt to design learning instruments that foster cognitive engagement and academic products. Strategically, this study emphasizes the importance of adopting AI-based learning tools to strengthen the quality, relevance, and global competitiveness of higher education.

Received 20 October 2025; Received in revised form 28 October 2025; Accepted 12 May 2026
Jurnal Eduscience (JES) Volume 13 No. 3 (2026)
Available online 30 June 2026

INTRODUCTION

The new paradigm of 21st-century education focuses on content mastery, the development of Higher-Order Thinking Skills (HOTS), including analytical, evaluative, and creative abilities, and the production of authentic outcomes in the form of tangible products that reflect high-level thinking processes. HOTS is seen as a fundamental competency for navigating the complexity of the digital age and technological disruption (Brookhart, 2023). In line with this, the outcome-based learning paradigm shifts the focus from "what lecturers teach" to "what students can produce" (Biggs & Tang, 2011), requiring learning instruments to facilitate procedural understanding and encourage students to produce work with academic and practical value. Recent research confirms that students engaged in learning that emphasizes HOTS and authentic products show significant improvements in problem-solving skills, creativity, and readiness to face global challenges (Heong et al., 2022; Latif et al., 2024). Thus, the integration of HOTS and outcomes in the modern learning framework can be a pedagogical strategy to increase the competitiveness of graduates at the global level

As an instructional tool commonly used in higher education, student worksheets (SW) serve to organize student activities in line with specific learning outcomes. However, recent research shows that many SW tools remain procedural, focusing on routine tasks or the repetition of concepts, thereby failing to facilitate the development of HOTS such as analysis, evaluation, and creation (Siregar, B. H., & Kairuddin, 2020). Based on relevant research, HOTS-based SW are valid and motivate students, but the activities designed have not optimally encouraged creativity and produced authentic academic products (Kahar et al., 2021). Another finding is that students have difficulty understanding material that remains rote and want SW that allows them to think critically and reflect on their own ideas (Shafira et al., 2022). Therefore, there is an urgent need to transform SW from a procedural to an outcome-based instrument that guides students in mastering the material and producing academic work and contextual solutions relevant to the real challenges of higher education.

In line with the need for learning instruments that improve HOTS and outcomes, the development of Artificial Intelligence (AI) in education presents significant opportunities for personalization, adaptation, and pedagogical innovation. AI enables personalized learning, real-time content adaptation, automated feedback, and more precise, data-driven support for understanding student needs (Holmes et al., 2021; Zawacki-Richter et al., 2019). This transformation shifts learning instruments from static to more dynamic, interactive, and responsive to student learning outcomes. However, a research gap remains, as most previous studies have emphasized the integration of AI in learning management systems (LMS) or intelligent tutoring systems (ITS), while studies on the application of AI in micro learning instruments such as student worksheets are relatively limited (Chen et al., 2020; Viberg et al., 2020). Therefore, integrating AI into SW is a strategy for improving HOTS and learning products. This is an alternative instructional innovation that can enhance the quality of learning in higher education.

The DEEPIR (Diagnosis, Exploration, Engagement, Product Development, Implementation, Reporting) model is presented as an innovative learning framework that follows a systematic process oriented towards improving HOTS and learning products (Siregar et al., 2025). This model enables students to go through the stages of initial diagnosis of learning needs, exploration of knowledge, active engagement in activities, development of real products, implementation of learning outcomes, and reflective reporting, all of which support the formation of comprehensive academic competencies. These characteristics align with the outcome-based education paradigm, which emphasizes the attainment of higher-order skills rather than merely procedural mastery (Biggs & Tang, 2011; Harden, 2020). The integration of AI into DEEPIR has the potential to strengthen each stage, for example, by enabling more accurate data-driven diagnosis, adaptive content exploration, more interactive engagement through intelligent systems, and reporting supported by automated feedback and personalization. Although the DEEPIR model has strong potential for integration into the design of 21st-century learning instruments, a research gap remains. There are no empirical studies that explicitly test the effectiveness of AI-integrated DEEPIR-based SW in higher education. This opens up significant research opportunities to develop new instructional models that can bridge the gap between higher-order thinking skills and the achievement of authentic academic products.

Student worksheets (SW) have long been used as an instructional tool, but most implementations remain procedural, making them less effective at encouraging students to produce tangible academic outcomes. This situation has created a significant gap, given that 21st-century higher education demands the creation of authentic products and the development of analytical, evaluative, and creative thinking skills. On the other hand, despite the widespread development of Artificial Intelligence (AI) in education, research has focused more on learning management systems and intelligent tutoring systems than on outcome-based instruments such as student worksheets (Zawacki-Richter et al., 2019; Holmes et al., 2021). As a result, opportunities to integrate AI into the DEEPIR model's systematic framework remain largely unexplored, despite its potential to produce adaptive, interactive, and productive learning tools. Therefore, there is an urgent need for a new framework that combines the strengths of the DEEPIR model with artificial intelligence to overcome the limitations of conventional SW and meet the demands of learning outcomes in the digital age.

Theoretically, this study contributes to enriching the literature on outcome-oriented SW development and to expanding the study of artificial intelligence (AI) integration into SW, which remains relatively unexplored in the context of higher education (Zawacki-Richter et al., 2019). From a practical perspective, this study offers an innovative model in the form of the AI-Integrated DEEPIR model, which can be used as a reference by lecturers in designing learning instruments that not only facilitate concept mastery but are also responsive to the needs of higher-order thinking skills (HOTS), active engagement, and the production of authentic academic work. Meanwhile, in the realm of policy, this research is relevant to the implementation of a competency-based higher education curriculum that emphasizes measurable learning outcomes and supports graduates' global competitiveness in the digital era (OECD, 2021; UNESCO, 2022). Thus, this study provides two contributions: (1) a conceptual contribution to the development of learning instrument theory, and (2) practical implications for classroom teaching and strategic direction for the formulation of higher education policy.

In light of the theoretical and practical gaps outlined above, this study aims to answer three main questions. First, to what extent does the AI-integrated DEEPIR model-based SW meet the criteria of validity, practicality, and effectiveness in the context of learning in higher education, considering that quality learning instruments must pass the test of these three aspects before being widely implemented (Plomp & Nieveen, 2019). Second, how this SW can facilitate students in producing authentic outcomes in the form of innovative learning media and scientific articles worthy of publication, in line with an outcome-based higher education orientation that emphasizes the relevance of academic products to real-world needs (Biggs & Tang, 2011). Third, how does this SW contribute to improving students' higher-order thinking skills (HOTS), particularly in analysis, evaluation, and creation, which are currently the main indicators of graduate quality in the 21st century (Brookhart, 2023; Phan et al., 2023). By answering these three questions, this research can enrich the academic discourse on outcome-based instrument design and provide an empirical basis for AI-based pedagogical innovation in higher education.

METHODOLOGY

Research Design

This study employed a design research approach by adapting the model proposed by Gravemeijer and Cobb (2006), which emphasizes the cyclical interaction between learning theory and practice. This model was adopted because it provides a framework that enables researchers to design, test, and refine learning instruments iteratively, thereby achieving optimal and comprehensive validity, practicality, and effectiveness. During the preparation and design stage, the researcher analyzed learning needs, identified gaps in conventional instruments, and designed student worksheets based on the DEEPIR model, integrating AI. The design experiment stage focused on implementing the SW prototype in the classroom to test the relationship between the design and the development objectives. Furthermore, during the retrospective analysis stage, researchers conducted both qualitative and quantitative analyses to assess the strengths and limitations of the SW. Then they revised the design based on the empirical findings and expert advice. The SW development procedure is outlined in Figure 1. Explain the research design of this study, including the reason for this design.

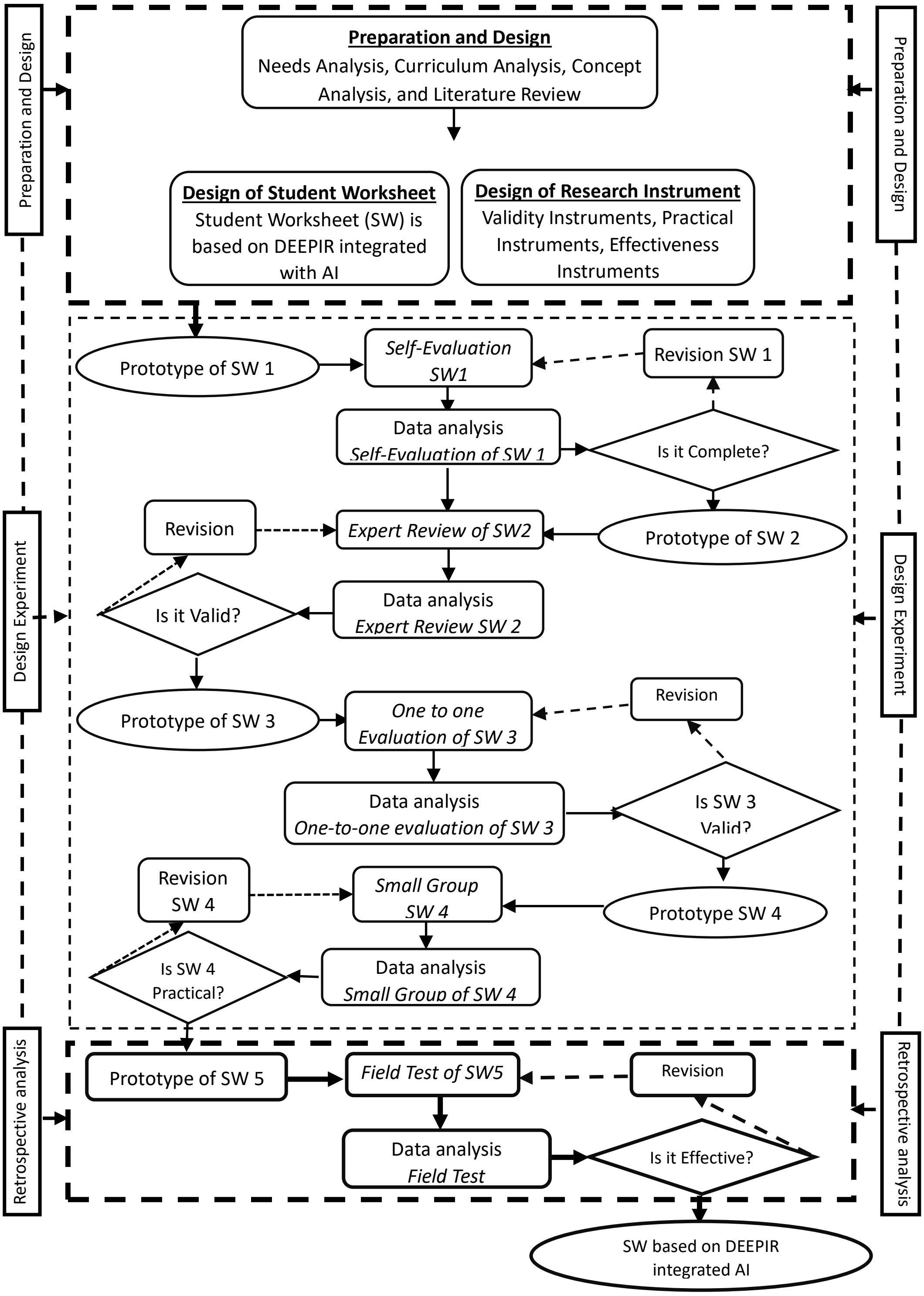

Figure 1. SW Development Procedure (Adapted from Gravemeijer and Cobb, 2006)

Participant

The study participants were 22 students from the Mathematics Education Study Program at Medan State University who were enrolled in the Learning Media course in the even semester of the 2024/2025 academic year. All participants were selected purposively because they already had pedagogical knowledge and initial experience in designing learning media, making them relevant for testing the effectiveness of Student Worksheets (SW) based on the DEEPIR model integrated with artificial intelligence. Participants played an active role in all stages of the research, from needs analysis and product design to implementation and reflective evaluation. Their direct involvement enabled researchers to obtain comprehensive empirical data on the validity, practicality, effectiveness, and appeal of the instrument in improving HOTS.

Data Collection

The data collected in this study is qualitative and quantitative data to obtain a comprehensive evaluation of AI-integrated DEEPIR-based student worksheets (SW). Qualitative data were obtained through interviews, observations of student interactions with the AI system, and notes taken during trials and implementation. This data was used to identify the strengths, limitations, and suggestions for development of the SW. Meanwhile, quantitative data included students' HOTS pretests and posttests, results from validity and practicality questionnaires, and students' responses to the SW.
Instrument

Data Analysis

Data analysis in this study was conducted both descriptively and inferentially to obtain a comprehensive overview of the quality of the developed SW. Descriptive analysis was employed to interpret data derived from learning observations, validity, practicality, effectiveness, and students' product quality (Nieveen, 1999). Meanwhile, the gain score concept (Meltzer, 2002) was used to assess the extent of improvement in HOTS following the implementation of AI-integrated DEEPIR-based worksheets. Each data component was analyzed through stages of reduction, presentation, and conclusion drawing to ensure the accuracy of the interpretation of the results. This approach enabled the researchers to comprehensively assess the interrelationships among the learning process, student engagement, and learning outcomes.

FINDINGS

Preparation and Design

The needs analysis revealed a mismatch between current instructional practices and outcome-based learning requirements that emphasize HOTS. The majority of SW used are procedural tasks, which do not encourage students to produce products through higher-order thinking processes, despite the curriculum listing achievements that require the ability to create media and scientific writing skills. The worksheets used do not facilitate the production of academic artifacts. Classroom observations indicate a low frequency of formative feedback and limited use of product-based rubrics. On the student side, there is variation in digital readiness and heterogeneous levels of initial HOTS abilities. At the same time, the infrastructure and lecturers' competence in using AI tools remain limited. These conditions require the design of instruments that are both adaptive and easy to operate. Critically, these findings emphasize that instructional interventions must simultaneously address task design to encourage students to produce authentic products through higher-order thinking processes.

The results of the curriculum analysis show a contradiction between the Course Learning Outcomes (CPMK) and the outcomes required by the course lecturers. Formally, the CPMK requires an increase in scientific communication, product creation, and product implementation skills in mini-research, but the learning plan and syllabus documents only emphasize cognitive aspects. The syllabus analysis also shows a lack of constructive alignment between learning objectives, learning activities, and assessment forms. For example, learning about media design is given a short assignment weight without a feedback-based revision mechanism, so students are not asked to iterate on the product until it meets academic standards. From a competency development perspective, this implies that simply inserting product assignments is not enough. Rather, it is necessary to develop valid rubrics, multi-level assignments that require synthesis and reflection, and to integrate peer review and publication activities into the assessment pathway. Therefore, the curricular recommendation that emerges is to reallocate the assessment weight towards process and product, and to strengthen HOTS-oriented rubrics.

Conceptual analysis maps out how each phase of DEEPIR (Diagnosis, Exploration, Engagement, Product Development, Implementation, Reporting) can be operationalized and strengthened by AI to produce truly outcome-based SW. In the Diagnosis phase, the AI analytic engine provides individual competency profiles so that tasks can be differentiated; in exploration, recommender systems and curated adaptive content expand the exploration path while maintaining cognitive complexity; the Engagement phase is supported by intelligent tutors and interactive scaffolding that maintain cognitive engagement; Product Development is enriched by AI-based authoring tools, such as suggestions and auto-rubric checking, so that students can design and revise media or article manuscripts; Implementation is useful for testing product feasibility in a real-world context in the form of mini-research; and Reporting encourages students to write down the results of product implementation in the form of scientific articles. In this context, the use of AI in the DEEPIR model is a strong framework. However, it requires clear implementation policies, ethical audits, and rubric design to ensure that SW transformation truly encourages HOTS improvement and produces quality academic outcomes.

The design stage begins with formulating clear and measurable operational learning objectives: (a) HOTS indicators at the analysis, evaluation, and creation levels; (b) product achievement criteria, namely learning media innovation and scientific articles with publication standards; and (c) activity indicators developed based on the DEEPIR model, d) AI as augmentation, not substitution. This framework ensures that each element of SW is designed to stimulate higher-order cognitive behavior and produce evidence in the form of academic artifacts.

Table 1. Functional Mapping of AI to the DEEPIR Phase in SW Development

DEEPIR Syntax	Student Activities	Integrated AI Function	Expected Output
Diagnosis	Identifying real problems in implementing mathematics learning media in partner schools.	Diagnostic engine, learning analytics, adaptive pretest to map competency gaps & real needs.	Profile of needs and recommendations for SW development paths.
Exploration	Explore theory, previous research, and good practices related to mathematics learning media.	Content recommender, AI summarizer, and semantic search to select relevant literature and practices.	Map of theories and good practices relevant to the development of learning media.
Engagement	Develop creative and innovative ideas for alternative learning media.	Idea generator based on engineering prompts, critical prompts to trigger evaluation of design alternatives.	Design innovative learning media that are creative, critical, and original.
Product Development	Producing learning media according to the design that has been prepared.	AI-assisted authoring tools (auto-layout, auto-check rubrics, interactive simulations, multimedia templates).	Learning media prototypes that comply with pedagogical and technical standards.
Implementation	Implement learning media in partner schools and analyze their effectiveness.	Analytics dashboard, user-interaction tracking, peer-AI review for data-based evaluation.	Media effectiveness data, analysis of student difficulties, and recommendations for improving learning media.
Reporting	Reflect, evaluate, and publish results in the form of scientific articles.	AI writing assistant (article structure, automatic citation, argumentation coherence, plagiarism checker).	Scientific articles/systematic reflections ready to be published in journals or proceedings

The integration of artificial intelligence (AI) into each phase of the DEEPIR model demonstrates its strategic role as a reinforcement of pedagogical and instructional functions in outcome-based learning. In the Diagnosis phase, AI supports adaptive pretesting and learning analytics to generate precise student competency profiles. In contrast, in the Exploration phase, content recommenders and adaptive prompts help tailor learning resources to students' cognitive levels, minimizing cognitive overload while still requiring supervision to ensure curriculum alignment. In the Engagement phase, intelligent chatbots and interactive scaffolding encourage collaborative and reflective participation, but the role of lecturers remains essential to prevent students from becoming dependent on automated systems. Furthermore, the Product Development phase is facilitated by AI through authoring tools, auto-checklists, and suggestion features, accelerating the production of articles or learning media. However, creativity and academic integrity must still be maintained. In the Implementation phase, field tests are conducted to ensure social and pedagogical acceptance. Finally, in the Reporting phase, AI functions as a tool for writing and evaluating products, but still requires expert interpretation to be pedagogically meaningful. Thus, AI functions as a pedagogical amplifier, strengthening the Diagnosis, Exploration, Engagement, Product Development, Implementation, and Reporting stages without replacing the important role of human experts in ensuring scientific quality and ethics.

The functional mapping of AI into the DEEPIR phases serves as a technical tool and as a reinforcement of pedagogical functions at each stage of learning. In the diagnosis phase, AI enhances the accuracy of learning needs mapping, enabling more targeted interventions. In contrast, during the exploration phase, AI accelerates access to, selection of, and synthesis of literature, ensuring students are well-trained to conduct evidence-based studies. Furthermore, in the engagement phase, AI serves as a facilitator of creative ideas while maintaining space for student originality. In the product development phase, AI becomes a partner in the production of learning media that meet pedagogical and technical standards. The implementation phase shows AI's contribution in generating empirical data through interaction-based analytics that strengthen product evaluation. In the reporting phase, AI supports the preparation of scientific articles by maintaining academic structure, ensuring argument coherence, and detecting plagiarism. Thus, this mapping confirms that AI is not a substitute for the roles of lecturers or students, but rather an accelerator of the cognitive, creative, and reflective processes needed to produce authentic products and high-quality scientific publications.

Design Experiment

The Development stage aims to produce a student worksheet (SW) prototype based on the DEEPIR model integrated with AI that meets the standards of content validity, operational practicality, and technical readiness for limited-scale testing. The basic principles of development are (a) fidelity to the established instructional design (constructive alignment with HOTS & outcome objectives), (b) technical modularity for ease of integration or iteration, (c) stakeholder involvement (lecturers, subject matter experts, technology experts) from the outset, and (d) ethical accountability.

In the initial prototype development stage, the development of SW based on the AI-integrated DEEPIR model focused on building an instrument framework that functionally integrates pedagogical flow with AI-based technology components. The process began with a prototype in the form of a conceptual design and module sketches to map the DEEPIR stages into learning activities, then continued to a digital-based prototype that allowed for the integration of AI components on a simple scale. The core components developed include a diagnostic engine for initial student competency profiling through adaptive pretesting, a rule-based content recommender to tailor learning resources to cognitive levels, an intelligent chatbot for interactive scaffolding, and AI-assisted authoring tools that support the drafting of academic products using auto-checklist formats. In addition, a simple reporting dashboard was developed to visualize student progress in real time. Critical analysis at this stage confirmed that the prototype served as an initial validation of conceptual and technical feasibility, emphasizing system modularity, AI explainability, and the balance between technological support and the role of lecturers as guarantors of academic quality.

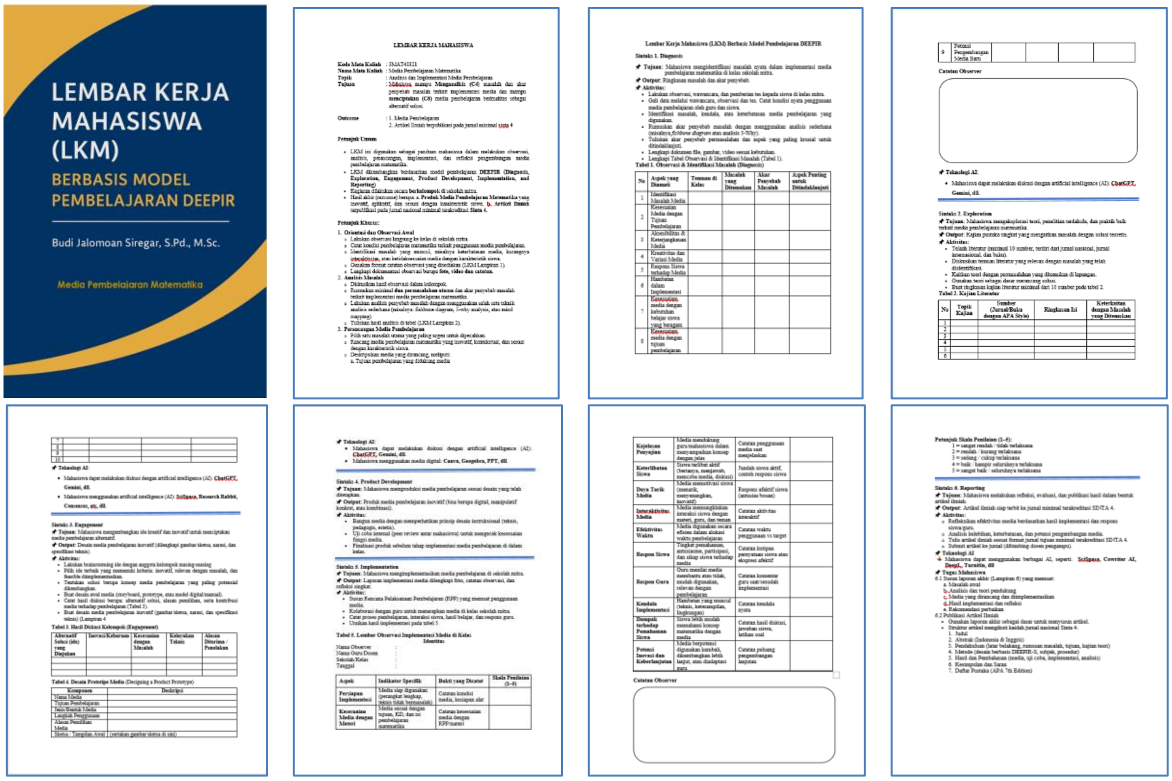

Figure 2. Prototype of DEEPIR-based SW Integrated with AI.

Self-Evaluation

The results of the self-evaluation show that the AI-integrated DEEPIR-based SW has good content coherence, activity flows aligned with DEEPIR syntax, and AI function integration that effectively supports diagnosis, exploration, and feedback. However, critically, there is still a need to enrich real cases in the diagnosis phase, expand learning resources in the exploration phase, and deepen interactive scaffolding in the engagement phase. These improvements are important to ensure the SW is structurally consistent, adaptive, and able to stimulate students' HOTS more deeply.

Expert Review

Expert reviews were conducted to assess the quality of SW using the DEEPIR model integrated with AI, focusing on three main aspects: content coherence, the suitability of the activity flow to DEEPIR syntax, and the integration of AI functions across each learning phase.

Table 2. Expert Review Results for DEEPIR-based SW Integrated with AI

Aspects	Findings of Strengths	Revision Notes	Recommendation
Content Coherence	The material aligns with the learning outcomes and integrates theory and practice in mathematics learning.	Some instructions are still general in scope, which could lead to multiple interpretations.	The task instructions need to be clarified to make them more focused and explicit.
Alignment of Activity Flow with the DEEPIR Syntax	The activity flow systematically follows the DEEPIR stages from diagnosis to reporting.	The transition between phases still lacks emphasis on the logical connection between student products.	Add bridging tasks to more integratively connect the outputs of each phase.
Integration of AI Functions	AI has supported needs analysis, resource exploration, interactive scaffolding, and auto-feedback.	AI functions at the reporting stage are still limited to simple analytics.	Develop AI features to provide scientific writing recommendations and evaluate article quality.

One-to-One Evaluation

The results of the One-to-One Evaluation show that SW based on the DEEPIR model integrated with AI has clear instructions, activity flow, and fairly good technology integration, but there are still several critical notes: (1) there are technical terms that require additional explanation, (2) transitions between phases are still challenging for students with low abilities, (3) AI integration has been proven to increase motivation and assist through resource recommendations and automatic feedback, although the responses are sometimes too generic and lack context, and (4) the cognitive level of some tasks is considered too high, requiring differentiation of scaffolding. These findings emphasize the need to improve instructional, technical, and content adaptation to optimize the effectiveness of SW.

To provide a more comprehensive picture of the practicality of AI-integrated DEEPIR-based SW, a small-group evaluation was conducted with lecturers and students as the main users. This evaluation focused on four aspects, namely usability, ease of use, time efficiency, and overall average score.

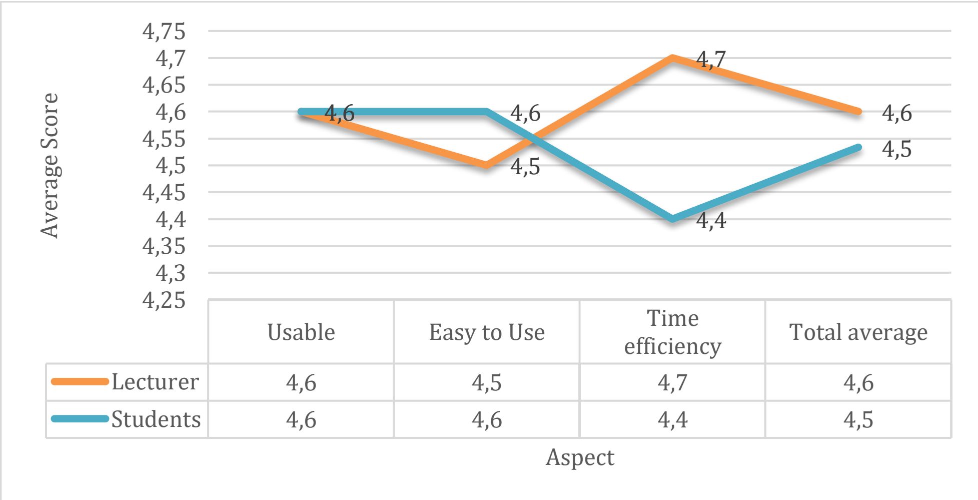

Figure 3. Expected Practicality Data of SW Based on The DEEPIR Model Integrated with AI

The results of the practicality evaluation show that the AI-integrated DEEPIR-based SW was rated very highly by both lecturers (average 4.6) and students (average 4.5). Both groups rated the SW's usability highly (4.6), indicating that it is easy to use and aligns with learning needs. Interestingly, students gave a higher score for ease of use (4.6) than lecturers (4.5), suggesting that students adapt more quickly to AI-based technology integration.

A significant difference emerged in terms of time efficiency, with lecturers rating SW as highly efficient (4.7), while students rated it lower (4.4). This indicates a perception gap: lecturers see the effectiveness of the designed activities, while students may feel a greater cognitive load and time demands. Thus, although SW is considered practical, improvements should focus on managing the workload to better balance perceptions of efficiency between lecturers and students.

Retrospective Analysis

The actual practicality of DEEPIR-based SW integrated with AI reflects the extent to which this tool is truly functional in a real learning context, particularly in terms of usability, ease of use, efficiency, and instructional relevance. The following graph presents the actual practicality results that form the basis for further critical analysis.

The practical results, as visualized in the graph, show that the DEEPIR-based SW integrated with AI received positive ratings from lecturers and students, with overall average scores of 4.6 and 4.5, respectively. The usability aspect received an equal score (4.6) from both groups, confirming that the SW is considered functional and suitable for learning needs in higher education. The ease of use aspect was also relatively consistent, with a score of 4.5 from lecturers and 4.6 from students, indicating that the instructions and structure of the SW were clear and did not cause significant technical obstacles. These findings support the practicality criteria as formulated by Nieveen (2007), that learning tools must be easy to use, relevant to needs, and able to support the teaching and learning process effectively.

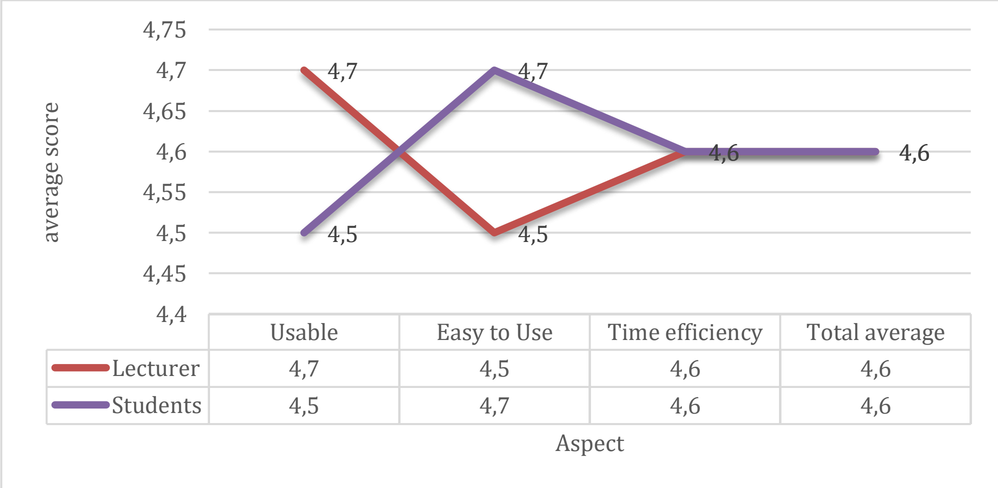

Figure 4. Actual Practicality Data of SW Based on The DEEPIR Model Integrated with AI

Based on the graph in Figure 4, the difference is evident in terms of time efficiency: lecturers give a very high rating (4.7), while students give a lower rating (4.4). This indicates a disparity in perception. Lecturers view SW as an instrument that accelerates the planning and evaluation of learning. At the same time, students still experience a relatively high cognitive load and time demands when completing HOTS-based tasks. This difference highlights the importance of strengthening the integration of AI functions, particularly to support time management and provide adaptive scaffolding, so the load on students is more proportional.

The effectiveness of SW is evaluated through two main indicators: improvements in higher-order thinking skills (HOTS) and students' positive responses to SW (attractiveness). The analysis focuses on the extent to which this instrument encourages the ability to analyze, evaluate, and create. In addition, attention is directed to the quality of students' learning experiences, as reflected in their perceptions of the attractiveness of SW innovations.

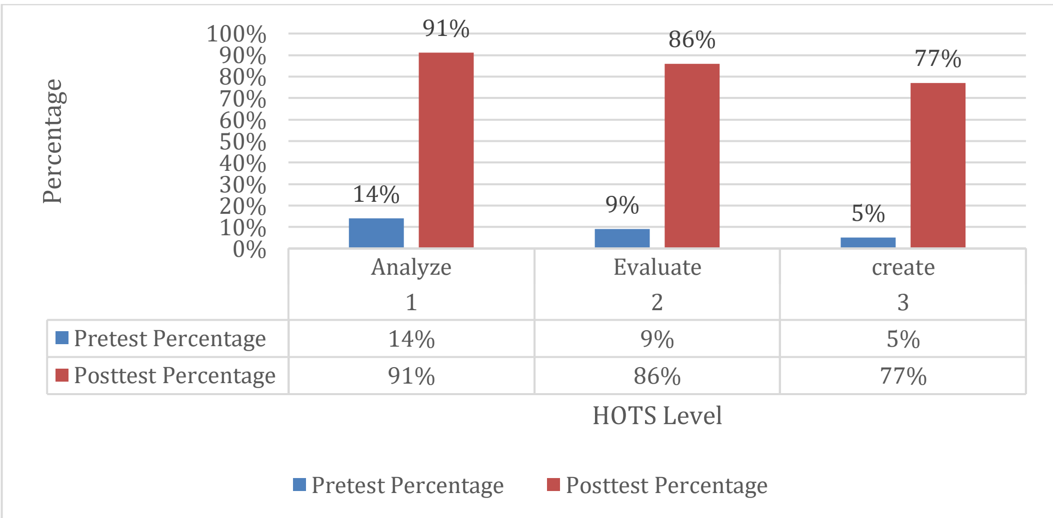

Figure 4. HOTS Achievement for Each Indicator

The data in Figure 5 reveal that students' HOTS achievement increased significantly across all indicators after using SW. In the pretest stage, student achievement was still extremely low, with only 14% in the analyze indicator, 9% in the evaluate indicator, and just 5% in the create indicator. However, after learning, there was a surge in achievement in the posttest, reaching 91% for analyze, 86% for evaluate, and 77% for create. This increase shows that the SW developed effectively encouraged students' HOTS, especially in analysis and evaluation. Although the achievement in creativity was relatively lower than the other indicators, this is understandable because creative skills require a more complex cognitive process. These findings indicate that SW not only serves as a learning tool but also facilitates the holistic development of students' critical and creative thinking skills.

Table 3. HOTS Achievement Based on N-gain Scores

N-Gain	Category	Number of Learners	Percentage	Average of gain
g > 0,7	high	17	77%	0.783
0,3 ≤ g ≤ 0,7	moderate	4	18%	
g < 0,3	low	1	5%	
Total		22	100%	

Based on the N-gain scores, the majority of students experienced an increase in HOTS in the high category. 17 of 22 students (77%) reached the high category, with a g value > 0.7, indicating a significant increase in HOTS mastery after learning using SW. Meanwhile, 4 students (18%) were in the moderate category (0.3 ≤ g ≤ 0.7), and only 1 student (5%) was in the low category (g < 0.3). The average N-gain of 0.783 indicates that SW is in the high effectiveness category, suggesting that its use can substantially improve students' higher-order thinking skills. These findings reinforce the evidence that AI-integrated DEEPIR-based SW is not only valid and practical but also effective in improving overall HOTS learning outcomes.

Table 4. Student Response Data to SW

Category	Number of students	Percentage	Number of Lecturers	Percentage
Positive Response	21	95%	1	100%
Negative Response	1	5%	0	0%
Not present	0	0%	0	0%
Total	22	100%	1	100%

The data in Table 4 show that the majority of students responded positively to SW (95%, or 21 out of 22 students), while only 1 student (5%) responded negatively, and no students did not respond. In addition, the lecturers involved gave a fully positive response (100%). These findings indicate that the SW has a high level of attractiveness according to Nieven's definition, namely the extent to which learning products can arouse interest, enjoyment, motivation, and foster a love of the learning process. This strong positive response confirms that SW not only functions as a learning tool but also creates an enjoyable and motivating learning experience, which ultimately serves as an important indicator of the effectiveness of learning products.

The outcomes achieved by students after implementing SW demonstrate their success in producing authentic, meaningful learning products. All students (100%) successfully created learning media. This demonstrates their ability to integrate conceptual knowledge with technological and pedagogical skills. The products produced not only function as learning aids, but also as concrete manifestations of creative, collaborative, and reflective thinking, which are the main characteristics of HOTS-based learning. In addition, student academic achievement is also reflected in the publication of scientific articles, as shown in Table 5.

Table 5. Number of Student Articles in Each Category

Category	Number of students			
	Draft	Summited	Published	Total
Article	0	4 (18%)	18 (82%)	22 (100%)

Table 5 shows the outcomes achieved by students after implementing SW, confirming its effectiveness in promoting academic performance through authentic outcomes. Of the total 22 students, 18 (82%) successfully published scientific articles in journals indexed at least in Sinta 4, 4 (18%) were at the submitted stage, and no students reached the draft stage. This data shows that almost all students achieved high productivity in producing scientific works worthy of publication, indicating the internalization of higher-order thinking skills, academic literacy, and reflective abilities in the context of research.

DISCUSSION

This study shows that DEEPIR-based SW integrated with AI is capable of fulfilling its main objectives: (1) encouraging students to produce products (media and scientific articles) and (2) improving students' HOTS. Quantitatively, the increase in HOTS was significant (n-gain = 0.783, high category), 100% of students produced learning media that met quality criteria, and 82% successfully published articles in journals accredited at least at Sinta 4. These findings are consistent with the hypothesis that the combination of a tiered instructional design (DEEPIR) and AI affordances (adaptive diagnosis, content recommendations, authoring assistance, and rapid feedback) can shorten the learning cycle, thereby strengthening the product-based learning process. The most contributing mechanisms appear to be related to (a) gradual scaffolding that forces transfer and synthesis, (b) faster formative feedback, which can accelerate product iteration, and (c) authoring features that reduce technical burdens so that students can focus on analytical and creative aspects. This is consistent with relevant research, where the combination of tiered instructional design with artificial intelligence affordances, such as adaptive diagnosis and authoring assistance mechanisms, has been empirically proven to shorten the learning cycle and improve higher-order thinking skills through customized scaffolding and rapid formative feedback (Contrino et al., 2024; Zuo et al., 2023). Furthermore, evidence shows that technology-enhanced feedback and research-designed scaffolding can encourage students to produce innovative products and achieve deep conceptual understanding, allowing authoring assistance to allocate cognitive resources to higher-level analysis and creation. (Kaldaras et al., 2024). These findings reinforce that the combination of tiered instructional design and AI affordances, such as adaptive diagnosis and authoring assistance, is an effective mechanism for shortening the learning cycle and strengthening HOTS outcomes.

Qualitatively, the results of expert evaluation and formative testing (through one-to-one and small group stages) confirm the coherence of the content and the suitability of the DEEPIR model flow, while also emphasizing the role of artificial intelligence (AI) as an augmentative agent that supports the process of diagnosis, exploration, and draft review; AI does not function as a substitute for human assessment, but rather as a support in the learning process. However, critical analysis of user data reveals two important facts. First, there is a difference in perceptions of time efficiency between lecturers (high scores) and students (lower scores), indicating that although the task design is considered efficient from a learning management perspective, the cognitive load required remains quite high for students, especially those with low digital readiness. Second, AI functions are not yet sufficiently sensitive to collaborative dynamics; the feedback generated tends to be individual-focused, thus not fully supporting coordination and evaluation of contributions in teamwork. These findings are in line with empirical evidence showing that adaptive scaffolding and high-tech feedback can reduce cognitive load while accelerating product iteration when designed for a collaborative context, but their effectiveness depends on the system's ability to provide analytics and scaffolding at the group level (Zuo et al., 2023; Kaldaras et al., 2024). Therefore, successful AI integration requires the development of features such as group analytics, team progress visualization, and scaffolding differentiation, as well as coaching in the form of training and task load readjustment so that technological support truly enhances collaborative processes and learning outcomes.

The theoretical contribution of this research is twofold. First, this research extends the DEEPIR instructional framework to student worksheets by demonstrating that artificial intelligence components can be systematically mapped to each DEEPIR syntax to strengthen pedagogical functions. In the diagnosis phase, artificial intelligence plays a role in providing personalization; in the exploration phase, AI supports the provision of curated evidence; in the engagement phase, its function lies in providing interactive scaffolding; in the product development phase, AI assists in the writing or authoring process; in the implementation phase, AI supports data-based evaluation; while in the reporting phase, AI facilitates analysis-based reflection.

Second, empirical evidence from product-based instrument development confirms these instruments' role as mediators between technology and higher-order thinking skills (HOTS). This shows that HOTS-oriented task design and assessment structure are as important as artificial intelligence functions in producing cognitive improvement. These results align with those of Zhuang et al. (2025), who emphasize that integrating AI into learning design should be grounded in a clear pedagogical syntax to ensure a connection between learning activities and higher-order cognitive outcomes. Similarly, Qian, Y. (2025) shows that the success of AI in learning depends not only on its technological sophistication, but also on the extent to which it can support task design that demands knowledge synthesis and creation. Thus, as emphasized by Ifenthaler et al. (2025), the effectiveness of AI in education lies in its integration with an instructional design framework that stimulates reflective reasoning and active knowledge construction.

Concerning practical implementation, these findings have several concrete implications. First, institutions need to allocate resources to develop curated content databases and reliable technical infrastructure. Second, strengthening faculty capabilities through training in dashboard analytics, peer review management, and substantive assessment is essential to maintaining the optimal role of humans in the loop. Third, SW design must incorporate differentiation mechanisms (adaptive prompts, scaffold tiers) and AI-based collaborative features (team contribution visualization, role recommendations) to reduce cognitive load and ensure inclusivity. Operationally, concrete recommendations include revising rubrics to be more explicit, enriching diagnostic cases, expanding the reference corpus for recommenders, and adding bridging tasks between DEEPIR phases.

The policy implications of this study indicate the urgency for universities to develop curricula that focus not only on cognitive aspects but also on authentic outcomes in the form of tangible products, such as learning media and scientific publications, as indicators of student achievement. To support this, institutions need to formulate ethical guidelines and governance for the use of artificial intelligence that cover aspects of transparency, attribution of use, and audit mechanisms for automated feedback. In addition, providing incentives in the form of funding for publications and allocating academic work time to lecturers and students who are active in scientific production is an important prerequisite. Policies structured in this way not only strengthen the sustainability of the model's implementation but also open opportunities for scalability across other study programs to increase academic competitiveness globally.

Another important aspect emerging from this study concerns the role of feedback mechanisms in strengthening students' self-regulated learning capacities within AI-assisted learning environments. The DEEPIR-based SW facilitated the completion of academic products and supported students in identifying conceptual weaknesses, evaluating information relevance, and revising their work through iterative cycles of diagnosis, reflection, and revision. AI-generated feedback provided rapid formative responses on scientific writing structure, media design, and conceptual accuracy, while lecturer feedback contextualized suggestions, clarified misconceptions, and deepened analytical reasoning. This finding indicates that the effectiveness of AI integration lies not merely in automation efficiency, but in the synergy between AI-supported feedback and human pedagogical guidance in fostering reflective and autonomous learning behaviors. In line with previous studies, AI-assisted environments can strengthen metacognitive monitoring and higher-order thinking when feedback is structured pedagogically and integrated with reflective learning activities (Ifenthaler et al., 2025). Therefore, future development of DEEPIR-based SW should incorporate more explainable feedback systems, adaptive reflection prompts, and differentiated scaffolding aligned with students' digital readiness and cognitive characteristics to ensure more meaningful and inclusive learning support.

This study has several limitations that must be openly acknowledged. First, the study focused only on one course, namely Learning Media, and on one cohort of students in the mathematics education study program (MESP 2023). This condition means that the results obtained cannot yet be generalized across disciplines or institutions. Second, although the research and development design provides strong evidence of initial feasibility and effectiveness, more rigorous causality testing still requires studies with a control design, for example, through a larger quasi-experimental design. Third, although the students' publication rate is relatively high, a more in-depth analysis is still needed to assess the editorial quality, the students' intellectual contributions, and the role of artificial intelligence in the writing process, to avoid ethical issues related to attribution. Fourth, this study has not explored the long-term impact, including HOTS retention and the transfer of skills into students' professional practice after graduation.

Based on these findings and limitations, the following research agenda is recommended: (1) replication studies in multi-departmental and cross-institutional contexts; (2) longitudinal studies to assess the sustainability of HOTS improvement and skill transfer; (3) controlled experiments to measure the specific effects of features; (4) an ethical-pedagogical evaluation of the use of generative AI in academic production; and (5) a cost-benefit study and readiness for implementation at institutions with limited resources.

In summary, this study demonstrates that combining DEEPIR instructional design with AI capabilities can transform STUDENT WORKSHEET from a procedural tool into an effective, outcome-oriented learning instrument that encourages HOTS and real academic production. Successful implementation requires technical refinements, such as collaborative AI features, explainability, policy investments in training, AI governance, and broader follow-up research to test generalization and long-term ethical implications.

CONCLUSION

This finding indicates that the development of Student Worksheets (SW) based on the DEEPIR model integrated with artificial intelligence has successfully transformed SW from procedural tools into outcome-oriented learning instruments that effectively encourage the production of academic products in the form of learning media and scientific articles and reflect an increase in students' HOTS; as empirically demonstrated by a significant increase in HOTS (n-gain = 0.783), 100% of students producing learning media, and 82% successfully publishing articles in journals accredited at least Sinta 4. These findings reinforce the theoretical claim that AI technology integration must be understood as pedagogical augmentation that functions optimally when contextualized in task designs that demand synthesis, evaluation, and creation and are supported by clear assessment rubrics; in other words, technology without proper instructional design will not produce substantial cognitive change. In practice, the AI-integrated DEEPIR model provides a replication framework for educators to design SW that balances scaffolding, product iteration, and authentic assessment; however, its implementation requires institutional investment in lecturer training, technical infrastructure, and AI governance policies (transparency, attribution, feedback audit). The limitations of this study include its single-cohort design, the absence of a control group, and the need to assess the long-term quality of publications and the transfer of competencies to professional practice. This calls for further research on a larger scale and a more rigorous experimental design. Thus, while these results are promising for improving the quality of higher education in the digital age, large-scale adoption must be accompanied by ethical safeguards and cost-benefit analyses to ensure sustainability and academic quality.

ACKNOWLEDGMENT

The LPPM Unimed supported this work through the PNBP Research Grant (Grant No. 0163/UN33.8/PPKM/PPT/2025, Year 2025).

REFERENCES

Biggs, J., & Tang, C. (2011). Teaching for quality learning at university (5th ed.). Open University Press.

Brookhart, S. M. (2023). How to assess higher-order thinking skills in your classroom (2nd ed.). ASCD.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. IEEE Access, 8, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Contrino, M. F., Reyes-Millán, M., Vázquez-Villegas, P., & Membrillo-Hernández, J. (2024). Using an adaptive learning tool to improve student performance and satisfaction in online and face-to-face education for a more personalized approach. Smart Learning Environments, 11(1), Article 6. https://doi.org/10.1186/s40561-024-00292-y

Gravemeijer, K., & Cobb, P. (2006). Design research from the learning design perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), Educational design research (pp. 17–51). Routledge.

Harden, R. M. (2020). Outcome-based education: The future is today. Medical Teacher, 42(8), 942–954. https://doi.org/10.1080/0142159X.2020.1776487

Heong, Y. M., Othman, W., Yunos, J. M., Kiong, T. T., Hassan, R., & Mohamad, M. M. (2022). The role of higher-order thinking skills in creative and critical problem solving. Thinking Skills and Creativity, 45, 101018. https://doi.org/10.1016/j.tsc.2022.101018

Holmes, W., Bialik, M., & Fadel, C. (2021). Artificial intelligence in education: Promises and implications for teaching and learning. Center for Curriculum Redesign. https://curriculumredesign.org/wp-content/uploads/AIED-Book-2021.pdf

Ifenthaler, D. (2025). Information processing re-visited: How artificial intelligence supports human-centered systems. AI-Enhanced Learning, 1(1), 229–243. https://doi.org/10.70725/465418kgasdx

Ifenthaler, D., Majumdar, R., Gorissen, P., et al. (2024). Artificial Intelligence in Education: Implications for Policymakers, Researchers, and Practitioners. Technology, Knowledge and Learning, 29, 1693-1710. https://doi.org/10.1007/s10758-024-09747-0

Kahar, M. S., Syahtputra, R., Bin Arsyad, R., Nursetiawan, N., & Mujiarto, M. (2021). Design of student worksheets focused on higher-order thinking skills (HOTS) in physics learning. Eurasian Journal of Educational Research, 96, 44–64. https://doi.org/10.14689/ejer.2021.96.3

Kaldaras, L., Wang, K. D., Nardo, J. E., Price, A., Perkins, K., Wieman, C., & Salehi, S. (2024). Employing technology-enhanced feedback and scaffolding to support the development of deep science understanding using computer simulations. International Journal of STEM Education, 11, Article (https://doi.org/10.1186/s40594-024-00490-7)

Latif, E., Zhou, Y., Guo, S., Shuchen, G., Lee, G., Liang, Z., … You, W. (2024). A systematic assessment of OpenAI o1-preview for higher-order thinking in education. arXiv.

Meltzer, D. E. (2002). The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores. American Journal of Physics, 70(12), 1259–1268. https://doi.org/10.1119/1.1514215

Nieveen, N. (2007). Formative evaluation in educational design research. In T. Plomp & N. Nieveen (Eds.), Educational design research (pp. 89–102). Enschede, The Netherlands: SLO.

Nieveen, N. (1999). Prototyping to reach product quality. In J. van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), Design approaches and tools in education and training (pp. 125–135). Springer. https://doi.org/10.1007/978-94-011-4255-7_10

OECD. (2021). OECD future of education and skills 2030: Curriculum analysis. OECD Publishing. https://www.oecd.org/education/2030-project/

Phan, H. P., Ngu, B. H., & Williams, A. (2023). Higher-order thinking, deep learning, and student achievement: A review of contemporary perspectives. Educational Psychology Review, 35(3), 79–104. https://doi.org/10.1007/s10648-023-09758-3

Plomp, T., & Nieveen, N. (Eds.). (2019). Educational design research: Introduction and illustrative cases. SLO.

Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. TechTrends. Advance online publication. https://doi.org/10.1007/s11528-025-01100-1

Shafira, F., Ristiono, R., Alberida, H., & Yogica, R. (2022). The need for high-order thinking skills-oriented student worksheets on the human circulatory system materials for grade XI students. Thinking Skills and Creativity Journal, 5(1), 17–21. https://ejournal.undiksha.ac.id/index.php/TSCJ/article/view/38500

Siregar, B. H., Amry, Z., Kairuddin, K., Nasution, F. M., Sagala, A. I. R., Hura, P., Daulay, A. P., & Gultom, S. A. (2025). Pengembangan model pembelajaran melalui outcome-based learning terintegrasi artificial intelligence untuk meningkatkan kemampuan berpikir tingkat tinggi [Unpublished research reports]. Universitas Negeri Medan.

Siregar, B. H., & Kairuddin. (2020). Development of students' activities sheet (SAS) of linear programming based on Krulik and Rudnick's problem-solving to improve problem-solving abilities. Journal of Physics: Conference Series, 1462(1), 012029. IOP Publishing. https://doi.org/10.1088/1742-6596/1462/1/012029

UNESCO. (2022). Reimagining our futures together: A new social contract for education. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000379707

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2020). The current landscape of learning analytics in higher education. Computers in Human Behavior, 89, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education, 16(39), 1–27. https://doi.org/10.1186/s41239-019-0171-0

Zhuang, M., Long, S., Martin, F., & Castellanos-Reyes, D. (2025). The affordances of Artificial Intelligence (AI) and ethical considerations across the instruction cycle: A systematic review of AI in online higher education. The Internet and Higher Education, 64, 101039. https://doi.org/10.1016/j.iheduc.2025.101039

Zuo, M., Kong, S., Ma, Y., Hu, Y., & Xiao, M. (2023). The effects of using scaffolding in online learning: A meta-analysis. Education Sciences, 13(7), 705. https://doi.org/10.3390/educsci13070705