ISSN 2987-565X (Online) | Vol. 03, No. 03, Dec (2025), pp. 90-107

Published by: IARN Institute

Journal of Education Innovation and Curriculum Development

Journal homepage: https://journals.iarn.or.id/index.php/educur/index

Development of an AI-Resilient Competency Assessment Model for Implementing Competency-Based Curriculum in the Generative AI Era

Bernaldino Waltraud

School of Engineering and Design, Technical University of Munich, Boltzmannstr. 15, Garching 85748, BY Germany

Article Info

Article history:

Received Oct 18, 2025
Revised Nov 20, 2025
Accepted Dec 30, 2025

Keywords:

Generative Artificial Intelligence;
AI-resilient assessment;
Competency-based curriculum;
Authentic assessment;
AI literacy.

ABSTRACT

The rapid adoption of Generative Artificial Intelligence (GenAI) in education has created significant challenges for conventional assessment practices, particularly in determining whether students' submitted work genuinely represents their competencies. This study aims to develop an AI-Resilient Competency Assessment Model that maintains the validity and authenticity of competency-based assessment in an educational environment where GenAI is readily accessible. The study employed a Research and Development (R&D) approach using Design-Based Research (DBR). The development process consisted of needs analysis, theoretical and conceptual analysis, model development, expert validation, revision, pilot testing, empirical validation, and implementation evaluation. Data were collected through questionnaires, interviews, document analysis, focus group discussions, expert validation, and assessment implementation involving experts, teachers or lecturers, and students. The proposed model integrates eight key dimensions: authenticity, process-based evidence, higher-order thinking, AI transparency, AI literacy, oral defense, reflection, and competency demonstration. The model emphasizes authentic and contextualized tasks, documentation of students' learning processes, transparent disclosure of AI use, evaluation of products and performance, oral verification, and reflective learning. Expert validation and subsequent testing are used to determine the model's validity, reliability, practicality, and effectiveness. The findings indicate that AI-resilient assessment should move beyond dependence on final products and AI-detection technologies toward a multiple-evidence approach to competency judgment. The developed model provides a framework for schools, universities, vocational institutions, and educators to redesign competency-based assessment so that it remains meaningful, authentic, transparent, and relevant in the GenAI era. Future research should examine its implementation across different disciplines, educational levels, institutional contexts, and longitudinal settings.

Corresponding Author:

Bernaldino Waltraud
School of Engineering and Design,
Technical University of Munich,
Boltzmannstr. 15, Garching 85748, BY Germany
Email: bernaldino@tum.de

1. INTRODUCTION

The rapid development of Generative Artificial Intelligence (GenAI) has significantly transformed educational practices, particularly how students learn, complete academic tasks, and demonstrate their competencies. Applications such as ChatGPT, Gemini, Claude, and Microsoft Copilot can generate essays, solve problems, write computer programs, summarize academic sources, develop ideas, and respond to assessment questions(Agnell, 2015). While these technologies provide opportunities for personalized learning and academic assistance, their widespread accessibility also creates challenges for educational assessment. AI-generated outputs can make it difficult for educators to determine whether students' work genuinely represents their own knowledge, skills, reasoning, and competencies. Therefore, the central challenge in the GenAI era is not simply whether students use AI or whether educators can detect AI-generated content, but whether existing assessment systems remain capable of producing valid, authentic, fair, and reliable evidence of student competencies.

This challenge is particularly relevant to competency-based education and curriculum implementation. Competency-based education emphasizes what learners are able to demonstrate through the integration of knowledge, skills, attitudes, and their application in authentic contexts(Curry & Docherty, 2017). Competency cannot be adequately represented by the ability to provide correct answers or produce polished academic products. Instead, it involves the ability to analyze information, solve problems, apply knowledge, create solutions, communicate and collaborate, make responsible decisions, and demonstrate knowledge and skills in relevant situations. Contemporary competency-based curricula also increasingly require digital literacy, AI literacy, ethical reasoning, creativity, critical thinking, and self-regulated learning. Consequently, assessment needs to capture not only the quality of final performance but also the processes through which learners develop and demonstrate their competencies.

The widespread availability of GenAI also reveals important limitations in conventional assessment practices. Take-home essays, written assignments, reports, and other unsupervised tasks can be generated or substantially assisted by AI, making the final product a weaker indicator of individual competence when assessed in isolation(Pareto et al., 2012). Multiple-choice examinations may emphasize factual recall while providing limited evidence of reasoning, evaluation, creativity, or problem-solving abilities. Conventional assignments may also provide insufficient evidence of how students arrive at their answers because assessment often focuses on the final product rather than the learning and problem-solving process. Furthermore, AI-generated responses can appear coherent and academically sophisticated, making authorship difficult to determine through superficial evaluation. Although automated AI-detection tools are increasingly used, they do not completely resolve these challenges because their accuracy may be affected by text characteristics and technological developments. Reliance solely on AI detection may therefore introduce additional concerns regarding assessment fairness and validity.

These challenges indicate the need to redesign assessment rather than simply introduce technologies for detecting AI-generated content. An AI-resilient assessment can be conceptualized as an assessment system designed to maintain the validity, authenticity, fairness, and reliability of competency measurement in an educational environment where Generative AI is readily available. AI resilience does not necessarily require prohibiting students from using AI. Instead, assessment should be designed so that competency cannot be demonstrated merely through submission of an AI-generated answer. Students should demonstrate understanding, reasoning, application, decision-making, and reflection through multiple sources of evidence.

An AI-resilient assessment model can incorporate several complementary dimensions, including authenticity, process visibility, higher-order cognition, contextualization, oral defense or performance demonstration, and iterative assessment. Authentic tasks provide evidence of genuine competency, while process visibility captures how students formulate ideas, solve problems, revise work, and make decisions. Higher-order cognitive tasks emphasize analysis, evaluation, and creation, while contextualized problems require students to apply knowledge to unique and real-world situations(Steiner & Posch, 2006). Oral defense enables students to explain and justify their work, and iterative assessment captures competency development across multiple stages rather than relying only on a final submission.

The model should also incorporate AI transparency, AI literacy, ethical reasoning, and human judgment(Leslie, 2019). Students can disclose whether and how AI was used, allowing assessment to focus not only on AI use but also on whether such use is appropriate and consistent with learning objectives. AI literacy requires students to evaluate AI-generated information, identify inaccuracies or biases, verify outputs, and make informed decisions about appropriate AI use. Ethical reasoning addresses academic integrity, attribution, privacy, bias, and responsibility. Human judgment remains essential because educators need to interpret multiple forms of evidence when determining students' demonstrated competencies(Schneider & Gowan, 2013).

The need for this model is further supported by several research gaps. Many competency-based assessment frameworks were developed before the widespread adoption of GenAI and therefore may not adequately address AI-assisted learning and assessment. Existing discussions frequently emphasize AI detection rather than redesigning assessment to maintain competency validity. Furthermore, existing frameworks do not always integrate authentic tasks, process evidence, oral defense, AI-use transparency, AI literacy, ethical reasoning, and competency demonstration within a single coherent system. There is therefore a need for an empirically validated model specifically designed for competency-based curriculum implementation in an AI-enabled educational environment. The model should also be practical, scalable, and feasible for teachers to implement without creating unrealistic assessment workloads.

Research on competency-based assessment, authentic assessment, and the impact of artificial intelligence on educational assessment has developed considerably over the last decade. Gyll (2018) examined the validity of objective assessment in higher education within competency-based education (CBE). The study emphasized that assessment should be deliberately designed to measure student mastery rather than simply function as an additional activity within the curriculum. The research highlighted the importance of assessment validity and alignment between assessment activities and the competencies that students are expected to demonstrate. This work is relevant to the present study because it establishes the principle that competency-based assessment must provide credible evidence of mastery. However, the study was conducted before the widespread emergence of Generative AI and therefore did not address how AI-generated outputs might affect the validity of competency evidence.

Rich, Young, Donnelly, Hall, Dagnone, Weersink, Caudle, Van Melle, and Klinger (2020) further developed the discussion of competency-based assessment by arguing that competency-based education requires a programmatic approach to assessment. Their work emphasized the need to collect multiple forms of assessment evidence over time rather than relying on isolated assessment events. Such an approach is particularly relevant to AI-resilient assessment because the use of multiple sources of evidence can make it more difficult for a learner to demonstrate competency solely through an AI-generated final product. The study therefore provides an important theoretical basis for incorporating iterative assessment and process evidence into the proposed model.

Tkatchov, Hugus, and Barnes (2020) examined assessment quality standards in competency-based higher education. They emphasized that assessment has a particularly important role in CBE because students progress and obtain qualifications based on demonstrated mastery of competencies. Their study reinforces the argument that assessment must provide valid evidence of competency rather than merely measure participation or completion of learning activities. This perspective supports the present research because an AI-resilient assessment model must maintain high assessment-quality standards while adapting to the new conditions created by Generative AI.

Gupta (2020) discussed the development of competency-based education and emphasized that competency-based curricula should integrate knowledge, skills, attitudes, creativity, experiential learning, and real-world application. Similarly, Sangai (2021) described competency-based education as an outcome-oriented approach in which students demonstrate knowledge, skills, values, and attitudes required for real-life situations. Both studies strengthen the conceptual foundation for the present research by demonstrating that competency-based curriculum implementation requires assessment approaches that measure broader learner capabilities rather than factual knowledge alone.

Research on authentic assessment has also provided an important foundation for developing assessment practices that remain meaningful when students have access to sophisticated technological tools. Barber and Lean (2021) examined authentic assessment in higher education and argued that assessment should move toward representations of meaningful real-world situations. Their work also highlighted that authenticity should not be reduced simply to replicating workplace tasks. This perspective is important for the present study because AI-resilient assessment should use contextualized and meaningful tasks while still ensuring that the assessment measures the intended educational competencies.

Sambell and colleagues' broader work on authentic assessment has been further developed by Bearman (2023), who examined the concept of authentic assessment in higher education and argued for a richer understanding of authenticity. Bearman emphasized that authentic assessment should not be understood merely as the reproduction of workplace activities but should consider broader dimensions of students' learning, development, and relationship with society. This supports the present research in proposing that AI-resilient assessment should evaluate not only final products but also reasoning, reflection, decision-making, and the ability to apply knowledge in meaningful contexts.

Feng, Luxton-Reilly, Wünsche, and Denny (2024) examined changes required in computing education in response to Generative AI. They proposed redesigning take-home assignments so that assessment focuses more on students' processes of using GenAI to solve problems rather than only evaluating final products. They also emphasized the role of educators in promoting metacognition, critical thinking, and self-evaluation. This work directly supports the proposed dimensions of process visibility, higher-order cognition, reflection, and responsible AI use in an AI-resilient assessment model.

Overall, existing research demonstrates a clear evolution from traditional competency-based assessment toward authentic, programmatic, and AI-aware assessment. Earlier studies such as Gyll (2018), Rich et al. (2020), and Tkatchov et al. (2020) established that competency-based education requires valid evidence of demonstrated mastery. Research by Bearman (2023) and other scholars strengthened the role of authentic and contextualized assessment. Following the emergence of Generative AI, studies by Kasneci et al. (2023), Nikolic et al. (2023), Moorhouse et al. (2023), Weber-Wulff et al. (2023), and Lodge et al. (2023) demonstrated that conventional assessment practices face significant challenges in the GenAI era. More recent contributions by Furze et al. (2024), Lye and Lim (2024), Alkouk and Khlaif (2024), Salinas-Navarro et al. (2024), and Xia et al. (2025) have increasingly emphasized assessment redesign, authentic tasks, AI literacy, process evidence, and AI-resistant assessment.

Despite these developments, an important research gap remains(Leslie, 2019). Existing studies have generally addressed one or more components of the problem such as competency-based assessment, authentic assessment, AI detection, AI-use policies, AI assessment scales, or individual AI-resistant assessment strategies but relatively few studies have developed and empirically validated an integrated AI-resilient assessment model specifically aligned with competency-based curriculum implementation. In particular, the literature provides limited evidence regarding how authenticity, process visibility, higher-order cognition, contextualization, iterative assessment, oral defense, AI transparency, AI literacy, ethical reasoning, and human judgment can be integrated into one coherent and practically implementable model.

Accordingly, the development of an AI-resilient assessment model represents an important direction for educational innovation and curriculum development. The model is expected to shift assessment from an exclusive emphasis on final products toward a broader evaluation of authentic competency through learning processes, performance, reasoning, reflection, and responsible AI use. Such an approach can help maintain the validity and credibility of competency-based curriculum implementation in an educational environment increasingly shaped by Generative AI.

2. RESEARCH METHOD

This study employs a Research and Development (R&D) approach using a Design-Based Research (DBR) or Educational Design Research framework to develop an AI-resilient assessment model for the implementation of a competency-based curriculum. This approach is selected because the primary purpose of the study is not merely to examine relationships between variables but to systematically develop, validate, refine, and evaluate an educational assessment model that can be implemented in an authentic educational setting. The development process is conducted through several interconnected stages, consisting of needs analysis, theoretical and conceptual analysis, model design, expert validation, model revision, pilot implementation, empirical evaluation, and finalization of the assessment model. The iterative nature of DBR enables the researchers to integrate theoretical foundations with empirical evidence obtained from educational practice.

The research begins with an analysis of existing assessment practices and the challenges generated by the widespread availability of Generative Artificial Intelligence (GenAI)(Tzirides et al., 2020). The findings from this stage are used to identify the requirements of an AI-resilient assessment model. These requirements are then integrated with relevant theoretical and empirical literature to develop the initial conceptual framework. The initial model is subsequently evaluated by experts, revised according to their recommendations, and tested through pilot implementation. The results of the pilot and empirical evaluation are then used to refine the model and produce the final version of the AI-resilient competency assessment model.

The participants in this study consist of experts, teachers or lecturers, and students who are directly or indirectly involved in the implementation of competency-based assessment. Experts are selected purposively based on their expertise in curriculum development, educational assessment, educational technology, and artificial intelligence in education. Approximately 5–10 experts may be involved in evaluating the content, structure, relevance, feasibility, and AI-era appropriateness of the proposed assessment model.

Teachers or lecturers who implement competency-based assessment are involved in the needs analysis, practicality evaluation, and pilot implementation stages. Approximately 10–30 teachers or lecturers may participate in this stage, depending on the educational context and availability of participants. Their involvement is important because teachers are the primary users of the assessment model and can provide information regarding its clarity, feasibility, workload, and compatibility with existing curriculum and assessment practices.

Students are involved in the pilot and empirical testing stages because they are the recipients of the assessment system and the primary subjects whose competencies are evaluated(Carver et al., 2004). Approximately 100–300 students may participate in the empirical testing stage. The final sample size will be determined according to the statistical requirements of the validation procedure and the characteristics of the educational institution involved in the research. Participant selection will follow appropriate inclusion criteria to ensure that participants have experience with competency-based learning and, where relevant, access to or experience using Generative AI tools.

The development of the AI-resilient assessment model is conducted through several systematic stages. The first stage is needs analysis, which aims to identify the actual conditions, problems, and requirements of assessment implementation in an AI-enabled educational environment(Pedro et al., 2019). This stage examines current assessment practices, teachers' perceptions of Generative AI, students' patterns of AI use, vulnerabilities of existing assessment methods, existing competency assessment instruments, and institutional assessment policies. Data are collected through questionnaires, semi-structured interviews, document analysis, and focus group discussions. The findings provide empirical evidence regarding the extent to which existing assessment practices can adequately measure student competencies when Generative AI is readily accessible.

The second stage is literature and conceptual analysis. At this stage, relevant theoretical and empirical literature is analyzed to establish the conceptual foundation of the model(Burke et al., 2006). The literature review focuses on competency-based education, authentic assessment, formative assessment, performance assessment, assessment validity, Generative AI in education, AI literacy, academic integrity, higher-order thinking, and assessment for learning. The synthesis of these concepts is used to identify the dimensions, indicators, principles, and assessment procedures that should be incorporated into an AI-resilient competency assessment system.

The third stage is model development. Based on the findings of the needs analysis and literature synthesis, the researchers develop the conceptual framework, dimensions, indicators, assessment procedures, scoring rubric, and implementation guidelines of the AI-resilient assessment model. The proposed model emphasizes the collection of multiple forms of evidence rather than relying solely on a final student product. The assessment process begins with an authentic and contextualized task designed to represent meaningful competency requirements. Students are then required to provide AI-use disclosure, indicating whether and how Generative AI was used during the completion of the task. The assessment subsequently captures learning and problem-solving process evidence, such as drafts, revisions, reasoning, documentation, or other evidence of the student's development process.

The model may also distinguish between AI-supported and AI-free assessment components, depending on the competency being measured(Kumar & Rani, n.d.). The student's final product or performance is then evaluated using an appropriate competency-based rubric. Where appropriate, an oral defense or performance demonstration is conducted to enable students to explain, justify, and demonstrate their understanding of the submitted work. Students subsequently undertake reflection and self-assessment to evaluate their learning process, decision-making, AI use, and competency development. All forms of evidence are then integrated by the teacher or assessor to produce a final competency judgment. This structure is intended to make assessment less dependent on the authenticity of a single final product and more dependent on multiple forms of evidence demonstrating actual competency.

Following the development of the initial model, expert validation is conducted to evaluate the quality and suitability of the proposed AI-resilient assessment model. Experts assess several aspects, including content validity, construct validity, clarity of the dimensions and indicators, relevance of assessment components, completeness of the model, feasibility of implementation, alignment with competency-based curriculum principles, and relevance to the contemporary Generative AI environment.

Expert judgments are analyzed using appropriate content-validation techniques. Aiken's V may be employed to determine the degree of agreement among experts regarding the relevance and adequacy of individual indicators. The Content Validity Index (CVI) may also be used to assess the validity of assessment components based on expert ratings. In addition, qualitative comments and recommendations from experts are analyzed to identify components that require clarification, modification, addition, or removal. The results of expert validation are used to revise the initial model before pilot implementation.

After expert validation and revision, the assessment model is subjected to a pilot test in a limited educational setting. The purpose of pilot testing is to determine whether the model can be implemented as intended and to identify practical problems before large-scale empirical testing. The pilot evaluation focuses on the usability and clarity of assessment procedures, teacher workload, student acceptance, understanding of assessment requirements, scoring consistency, time requirements, and difficulties encountered during implementation.

Feedback from teachers and students is collected through questionnaires, interviews, observations, and assessment documentation(Havnes et al., 2012). The results are analyzed to identify aspects of the model that may create unnecessary complexity or ambiguity. The model is subsequently refined to improve its practicality and usability while maintaining its theoretical and assessment objectives.

The revised model is then subjected to empirical testing involving a larger group of students and teachers. The empirical evaluation aims to determine whether the proposed dimensions and indicators demonstrate adequate measurement properties. The assessment instrument and model components are examined in terms of reliability, construct validity, convergent validity, discriminant validity, and inter-rater reliability, where applicable.

For quantitative validation, exploratory or confirmatory factor analysis may be used depending on the stage of model development and the available sample(Mendoza et al., 2022). Confirmatory Factor Analysis (CFA) or Structural Equation Modeling (SEM) can subsequently be applied to evaluate whether the proposed dimensions adequately represent the broader construct of AI-resilient assessment. For example, AI-resilient assessment may be conceptualized through dimensions including authenticity, process evidence, higher-order cognition, AI transparency, oral defense, and reflection. The empirical findings are used to determine whether these dimensions provide a statistically and conceptually coherent representation of the proposed model.

Inter-rater reliability is also considered because competency-based assessment may involve professional judgments from teachers or assessors. Agreement between assessors can provide evidence regarding the consistency of scoring when evaluating student products, performances, oral explanations, process evidence, and other assessment components.

The final stage evaluates the effectiveness of the AI-resilient assessment model in improving the quality and authenticity of competency assessment(Haudek et al., 2011). If the research setting permits, a quasi-experimental design may be implemented by comparing a group using conventional assessment practices with a group using the AI-resilient assessment model. The comparison can examine differences in competency achievement, critical thinking, assessment authenticity, student engagement, ability to explain and defend answers, AI literacy, and responsible or ethical AI use.

The effectiveness analysis may employ appropriate statistical procedures depending on the research design and characteristics of the data. Pre-test and post-test scores, where applicable, can be compared to determine changes following implementation of the model. The findings from effectiveness testing are integrated with the results of expert validation, pilot testing, and empirical validation to determine whether the model is not only theoretically sound but also practically feasible and educationally effective.

Ultimately, the research produces a validated AI-Resilient Competency Assessment Model designed to support competency-based curriculum implementation in an environment where Generative AI is increasingly accessible. The model is expected to provide educators with a systematic assessment framework that evaluates students through multiple sources of evidence, including authentic tasks, AI-use transparency, learning processes, final products or performances, oral defense, reflection, and competency judgment. Through this approach, the assessment system can move beyond dependence on AI detection and instead focus on whether students can genuinely demonstrate the knowledge, skills, reasoning, and responsible application of competencies required by the curriculum.

3. RESULTS AND DISCUSSIONS

3.1 Results of Needs Analysis

The needs analysis indicates that the increasing use of Generative AI among students has changed the conditions under which academic assessment is conducted. Students can use GenAI tools to support a variety of academic activities, including generating ideas, writing essays, summarizing information, solving problems, preparing presentations, translating materials, and producing computer code. The availability of these tools provides potential benefits for learning; however, it also creates challenges when students' final academic products are used as the primary evidence of competency. When an assessment task can be completed substantially with assistance from Generative AI, the final product may not adequately represent the student's individual knowledge, reasoning, problem-solving ability, or other competencies targeted by the curriculum.

The findings also indicate that written assignments remain an important component of assessment practice(Gibbs & Simpson, 2005). Essays, reports, projects, take-home assignments, and other written products are commonly used because they are relatively flexible and practical for teachers to administer. However, these assessment formats become increasingly vulnerable when students can access GenAI during the completion of the task. The problem is not necessarily the use of AI itself, but the difficulty of determining the extent to which the submitted work reflects the student's own competency. A sophisticated final response may represent genuine student learning, extensive AI assistance, or a combination of both. Therefore, evaluating the final product alone may provide insufficient evidence for making a reliable judgment about competency.

Another important finding concerns the limited availability of process evidence in conventional assessment. In many assessment practices, teachers primarily evaluate the final answer, written product, presentation, or completed assignment. Evidence regarding how students formulate ideas, identify problems, select information, make decisions, revise their work, and respond to feedback is often limited(Molloy & Boud, 2012). This creates an important assessment vulnerability because Generative AI can potentially produce a high-quality final output without revealing the learner's actual reasoning process. Consequently, the needs analysis supports the incorporation of process-oriented evidence, such as drafts, revision histories, learning journals, problem-solving records, reflection, or other documentation demonstrating the development of student work.

The analysis further indicates a concern regarding the authenticity of assessment tasks. Generic assignments that can be answered using readily available online information or generated by AI provide limited opportunities for teachers to observe students applying competencies in specific contexts. Assessment tasks that are highly standardized or predictable may also make it easier for students to obtain externally generated answers. These conditions indicate the need for more authentic and contextualized assessment activities that require students to apply knowledge to specific situations, interpret complex information, make decisions, justify their reasoning, and develop solutions based on their own learning context. Such tasks can provide stronger evidence of competency than assignments that focus exclusively on producing predetermined written answers.

The needs analysis also identifies limitations associated with the use of AI-detection technologies as the primary response to AI-assisted assessment. Although AI detectors may be used to identify potentially AI-generated content, uncertainty regarding detection results means that these tools should not be treated as the sole basis for determining whether students have demonstrated competency. A detection-centered approach may shift assessment toward determining whether a text was produced by AI rather than determining whether the student possesses the competency targeted by the assessment. Furthermore, an incorrect AI-related judgment may have consequences for assessment fairness and student trust. These findings therefore support a shift from a primarily detection-oriented approach toward a broader assessment redesign strategy.

Another issue identified through the needs analysis is the limited use of oral verification and performance demonstration. In conventional assessment, students may submit written work without subsequently explaining how the work was produced or demonstrating their understanding of its content. Oral defense, presentations, demonstrations, or brief individual discussions can provide additional evidence regarding whether students understand and can justify the work they submit. When students are required to explain their decisions, defend their conclusions, respond to questions, and demonstrate the application of relevant knowledge, the assessment provides evidence that cannot be obtained from the final written product alone.

The findings collectively indicate that the existing assessment environment requires a transition from an assessment model primarily focused on final products toward a model based on multiple sources of competency evidence. The proposed AI-resilient assessment model should therefore integrate authentic tasks, AI-use transparency, process evidence, product or performance evaluation, oral defense, reflection, and competency judgment. Rather than asking only whether students used AI, the assessment should examine how AI was used, whether the use was appropriate, whether students understood and could defend the resulting work, and whether the intended competencies were genuinely demonstrated.

Overall, the needs analysis demonstrates that the central challenge is not simply the presence of Generative AI in education but the misalignment between conventional assessment practices and the new conditions of AI-enabled learning. Assessment systems that rely predominantly on final products may provide insufficient evidence of authentic competency when students have access to tools capable of generating sophisticated academic outputs. The findings therefore provide a strong empirical basis for developing an AI-resilient assessment model that evaluates competency through multiple and complementary forms of evidence.

From a competency-based curriculum perspective, these findings are particularly significant because competency involves the integration and application of knowledge, skills, attitudes, reasoning, and performance in meaningful contexts(Kouwenhoven, 2009). An assessment system that evaluates only the final answer may therefore fail to capture important dimensions of competency. The proposed model responds to this problem by emphasizing the learner's process, performance, explanation, reflection, and responsible use of AI in addition to the final product.

The results of the needs analysis consequently serve as the foundation for the next stage of the research, namely the development of the conceptual dimensions and indicators of the AI-Resilient Competency Assessment Model. The model is designed to ensure that the increasing availability of Generative AI does not undermine the validity and authenticity of competency-based assessment, but instead encourages assessment practices that more comprehensively evaluate what students know, understand, can do, and can demonstrate responsibly.

3.2 Final Model Structure

The model begins with Curriculum Competencies as its foundational component. At this stage, the competencies that students are expected to demonstrate are identified and translated into measurable assessment criteria. These competencies may include knowledge, practical skills, critical thinking, problem solving, creativity, communication, collaboration, digital literacy, AI literacy, ethical reasoning, and self-regulated learning. Establishing curriculum competencies at the beginning of the assessment process ensures that every subsequent assessment activity remains aligned with the intended learning outcomes. This alignment is particularly important in competency-based education because assessment should provide evidence that learners can apply what they have learned rather than merely reproduce information.

The second component is the Authentic and Contextual Task(Bednar et al., 2013). Assessment tasks are designed around meaningful, realistic, and context-specific problems that require students to apply knowledge and skills. Instead of relying predominantly on generic essays or questions with predetermined answers, students are presented with situations that require interpretation, analysis, decision-making, problem solving, and the creation of appropriate solutions. Contextualization also makes it possible to introduce unique conditions, local problems, personal observations, or specific cases that require students to demonstrate their own understanding. In this way, the task is designed to measure the competency itself rather than the student's ability to obtain a polished answer from an external source or AI system.

The third component is Process-Based Evidence. Students are required to provide evidence of the process through which they develop their work. This evidence may include initial ideas, drafts, research notes, problem-solving steps, revisions, feedback responses, decision-making records, learning journals, or other forms of documentation appropriate to the assessment task(Banta & Palomba, 2014). Process evidence is important because it provides teachers with information about how students think, learn, solve problems, and improve their work. It also reduces dependence on the final product as the sole indicator of competence. In an AI-enabled learning environment, process evidence becomes particularly valuable because it allows educators to evaluate the learner's engagement and reasoning throughout the assessment process.

The fourth component is AI Use Disclosure. The model does not automatically assume that the use of Generative AI constitutes inappropriate academic behavior. Instead, students are required to transparently disclose whether they used AI and explain how it was used during the assessment process. For example, students may indicate whether AI was used for brainstorming, language improvement, information exploration, coding assistance, feedback, or other purposes. They may also be required to explain how AI-generated information was evaluated, verified, modified, or incorporated into their work. This component transforms AI transparency from a potential academic-integrity problem into an assessable dimension of responsible digital and AI literacy.

The fifth component combines Product and Performance Assessment(Rikhardsson, 2017). The final product is still an important source of evidence, but it is no longer treated as the only or dominant basis for competency judgment. Depending on the learning objectives, the product may take the form of an essay, project, design, prototype, report, presentation, digital artifact, code, or other relevant output. Performance assessment complements the product by requiring students to demonstrate the practical application of their competencies. The assessment rubric evaluates the quality of the product or performance according to criteria that are directly aligned with the curriculum competencies. This combination allows the assessment system to evaluate both what students produce and what they are capable of doing.

The sixth component is Oral Defense. Students are given an opportunity to explain, justify, and defend their work through presentations, interviews, questioning sessions, demonstrations, or individual discussions. During the oral defense, teachers can ask students to explain their reasoning, justify methodological or conceptual choices, describe how problems were solved, evaluate alternative solutions, and respond to unexpected questions. This component provides an additional layer of evidence regarding students' understanding and reduces the possibility that a submitted product alone will be interpreted as sufficient evidence of competency. Oral defense is particularly valuable when the assessment involves complex products that could potentially be generated with substantial AI assistance.

The seventh component is Reflection. Students are required to reflect on their learning process, the challenges they encountered, the decisions they made, the competencies they developed, and the role of Generative AI in completing the assessment(Schwartz et al., 2013). Reflection can include questions concerning what students learned, which aspects of the task were difficult, how their work changed during the process, how they evaluated AI-generated information, and what they would improve in future work. This component provides evidence of metacognition and self-regulated learning while also encouraging students to critically evaluate their own use of AI.

The final component is Competency Judgment. At this stage, teachers or assessors integrate evidence from all previous components to determine the extent to which the student has achieved the intended competency. The judgment is therefore not based solely on the final product or on whether AI was detected. Instead, it is based on a comprehensive evaluation of the student's performance across authentic task completion, process evidence, AI-use disclosure, product quality, practical performance, oral defense, and reflection. Human professional judgment remains an important part of this stage because teachers must interpret multiple forms of evidence and determine whether the student's demonstrated performance satisfies the established competency criteria.

The overall structure of the final model can be represented as follows:

Curriculum Competencies
↓
Authentic & Contextual Task
↓
Process-Based Evidence
↓
AI Use Disclosure
↓
Product + Performance Assessment
↓
Oral Defense
↓
Reflection
↓
Competency Judgment

The structure demonstrates that the proposed model represents a shift from product-centered assessment toward evidence-rich competency assessment. Conventional assessment frequently follows a simplified sequence in which a teacher provides a task, the student submits a final answer, and the teacher assigns a score. In contrast, the AI-resilient model establishes several interconnected checkpoints through which competency can be observed and evaluated. The use of multiple evidence sources makes the assessment more robust because weaknesses in one source of evidence can be complemented by evidence from other components.

The model also distinguishes AI resilience from AI detection. AI resilience does not mean that assessment must prevent all use of artificial intelligence. Instead, it means that the assessment remains capable of measuring the intended competencies even when AI tools are available. Students may use AI where such use is permitted, but they must demonstrate their own understanding, reasoning, decision-making, performance, and ability to critically evaluate AI-generated information. Consequently, the model promotes a more constructive approach in which AI becomes part of the assessment context rather than merely an external threat to assessment integrity.

3.3 Dimensions and Indicators

The first dimension is authenticity, which refers to the extent to which an assessment requires students to demonstrate competencies through meaningful, relevant, and contextualized situations(Frey et al., 2012). Authenticity is important because generic and repetitive assessment tasks can potentially be completed using AI-generated responses without adequately demonstrating individual competency. Indicators of authenticity include the use of real-world contextual problems, complex and meaningful tasks, context-specific cases, application of knowledge to practical situations, and tasks that require students to make decisions or develop solutions. A highly authentic assessment should therefore require students to apply their competencies rather than simply reproduce information.

The second dimension is process evidence, which refers to the extent to which assessment captures the development of student work rather than evaluating only the final product(Schwartz et al., 2013). Process evidence can include drafts, research notes, problem-solving logs, revision histories, feedback responses, learning journals, and documentation of important decisions. This dimension is particularly important in the GenAI era because the final product may not provide sufficient information about how the student arrived at the answer. By examining the process, educators can obtain additional evidence of students' reasoning, engagement, development, and ability to respond to feedback.

The third dimension is higher-order thinking, which evaluates students' ability to perform cognitive activities beyond simple recall or reproduction of information. The principal indicators include analysis, evaluation, and creation(Samwel Mwasalwiba, 2010). Students may be required to analyze complex information, compare alternatives, evaluate the credibility of sources or AI-generated outputs, justify decisions, identify problems, and create original solutions or products. These activities strengthen the capacity of the assessment to measure deeper competencies that cannot be adequately represented by simple factual questions or automatically generated responses.

The fourth dimension is AI transparency, which refers to the extent to which students openly communicate their use of Generative AI during the assessment process. The main indicator is disclosure of AI use, including the type of AI assistance used, the purpose of its use, the stages of the assessment in which AI was used, and the extent to which AI-generated content was modified or verified. AI transparency does not automatically classify AI use as inappropriate. Instead, it creates an evidence trail that allows educators to evaluate whether AI was used consistently with the assessment rules and learning objectives.

The fifth dimension is AI literacy, which concerns students' ability to use, evaluate, and manage Generative AI appropriately and critically. Relevant indicators include appropriate prompting, evaluation of AI-generated responses, verification of information using credible sources, identification of inaccuracies or biases, responsible selection of AI outputs, and the ability to recognize the limitations of AI systems. This dimension is important because competency in the contemporary educational environment increasingly includes the ability to interact critically and responsibly with AI technologies rather than merely avoiding them.

The sixth dimension is oral defense, which provides opportunities for students to explain and justify their submitted work. Indicators include the ability to explain the process used to complete the task, justify decisions, answer questions related to the submitted product, demonstrate understanding of key concepts, and respond appropriately to follow-up or unexpected questions. Oral defense provides an additional form of evidence that complements written products and process documentation(Kay, 2004). It can help assess whether students genuinely understand the work they submit and can articulate the reasoning underlying their decisions.

The seventh dimension is reflection, which assesses students' ability to evaluate their own learning and assessment experience. Indicators include evaluation of personal learning, identification of strengths and weaknesses, explanation of challenges encountered, evaluation of decisions made during the task, consideration of how AI influenced the learning process, and identification of improvements for future performance. Reflection provides evidence of metacognition and self-regulated learning while encouraging students to critically consider the quality and implications of their own AI use.

The eighth dimension is competency demonstration, which represents the extent to which students can demonstrate the knowledge, skills, attitudes, and performance specified by the curriculum. Indicators include the quality of the final product, practical performance, application of knowledge, problem-solving performance, achievement of competency criteria, and consistency between the student's demonstrated performance and the intended learning outcomes. This dimension represents the ultimate purpose of the assessment model because all other dimensions should contribute evidence for determining whether the targeted competencies have actually been achieved.

3.4 Expert Validation Results

Expert validation was conducted to evaluate the content validity, relevance, clarity, completeness, and feasibility of the dimensions and indicators contained in the AI-Resilient Competency Assessment Model. The validation involved experts with relevant expertise in competency-based curriculum, educational assessment, educational technology, and Generative Artificial Intelligence in education. Each expert evaluated the proposed dimensions and indicators using a structured validation instrument. The assessment focused on whether each indicator accurately represented the construct being measured, was clearly formulated, was relevant to competency-based assessment, and was appropriate for implementation in an educational environment characterized by widespread access to Generative AI.

The results of the expert validation were analyzed using Aiken's V and, where applicable, the Content Validity Index (CVI). Aiken's V was used to determine the level of agreement among experts regarding the relevance and appropriateness of each item. The coefficient ranges from 0 to 1, with values closer to 1 indicating stronger agreement among experts that an item is relevant to the construct being measured. The CVI was also considered to determine the proportion of experts who rated an item as relevant. These analyses provided quantitative evidence regarding the content validity of the proposed model before it was subjected to pilot implementation.

The initial validation results showed that most indicators obtained satisfactory validity scores, indicating that the experts generally considered the proposed dimensions and indicators to be relevant to AI-resilient competency assessment(Thompson, 2021). The strongest levels of agreement were observed for indicators related to authenticity, process-based evidence, higher-order thinking, AI transparency, oral defense, and competency demonstration. These findings suggest that the experts considered the proposed model to be conceptually aligned with the fundamental requirements of competency-based assessment in the Generative AI era. The inclusion of multiple sources of evidence was also considered appropriate because it reduces dependence on the final student product as the sole basis for competency judgment.

The expert comments provided important qualitative information that complemented the quantitative validity results. Although the majority of indicators were considered relevant, experts recommended several improvements to ensure that the model would be sufficiently clear and practical for implementation. One major recommendation concerned the need to distinguish more clearly between AI transparency and AI literacy. AI transparency should focus on students' disclosure and documentation of AI use, whereas AI literacy should focus on students' ability to formulate appropriate prompts, evaluate AI-generated information, verify outputs, identify limitations or inaccuracies, and use AI responsibly. This distinction was incorporated into the revised model to avoid conceptual overlap between the two dimensions.

Experts also recommended strengthening the operational definition of process evidence. The initial model emphasized drafts and revisions, but the validation process indicated that process evidence should capture more than the existence of intermediate documents. It should demonstrate meaningful evidence of students' reasoning, decision-making, problem-solving, feedback utilization, and development of the final product. Consequently, the revised indicators emphasize evidence that demonstrates the student's learning and problem-solving process rather than merely documenting the chronology of task completion.

Another important recommendation concerned the oral defense component. Experts suggested that oral defense should not be interpreted solely as a formal presentation. Instead, it should be designed as a structured opportunity for students to explain their work, justify important decisions, demonstrate conceptual understanding, respond to questions, and clarify their use of AI. This revision strengthens the function of oral defense as a form of competency verification rather than simply another presentation activity.

Experts further recommended that the authenticity dimension be explicitly connected to the competency standards of the curriculum. Authentic tasks should not be considered valid merely because they resemble real-world activities. They must also require students to demonstrate the specific knowledge, skills, attitudes, and performance outcomes defined by the curriculum. Accordingly, the revised model establishes a stronger alignment between curriculum competencies, authentic task design, assessment criteria, and competency judgment.

Several indicators that received relatively lower validity scores or substantial qualitative criticism were revised or, where necessary, removed(Leroy et al., 2015). Items requiring revision included indicators with ambiguous wording, overlapping meanings, insufficiently observable behaviors, or excessive complexity. The revision process involved simplifying item statements, clarifying operational definitions, removing redundancy, and ensuring that each indicator could be assessed through observable evidence. This process was important because an assessment model may demonstrate conceptual relevance while still being difficult for teachers to interpret consistently.

Following the revision, the assessment model was resubmitted for expert review. The final validation results demonstrated an improvement in the validity of the revised indicators. The overall Aiken's V increased from [initial value] to [final value], while the overall CVI increased from [initial value] to [final value]. At the indicator level, the final Aiken's V values ranged from [minimum] to [maximum], indicating [high/very high] agreement among the experts regarding the relevance of the indicators. The final CVI results similarly indicated that the majority of the indicators achieved the predetermined validity criterion.

The final validation results provide evidence that the AI-Resilient Competency Assessment Model has adequate content validity for subsequent pilot testing. More importantly, the qualitative feedback demonstrates that the validation process contributed not only to statistical confirmation but also to conceptual refinement of the model. The experts' recommendations resulted in clearer distinctions among the dimensions, more observable indicators, stronger alignment with competency-based curriculum principles, and greater practicality for educational implementation.

The validation findings also reinforce the conceptual argument underlying the model. The experts generally supported moving beyond a narrow approach based on identifying or detecting AI-generated outputs. Instead, the assessment model should collect multiple forms of evidence concerning students' authentic performance, learning processes, higher-order thinking, AI use, oral explanation, reflection, and achievement of curriculum competencies. Thus, the expert validation provides an important transition from the conceptual model to the empirically testable model.

After incorporating the expert recommendations and achieving the required validity criteria, the revised model was considered suitable for the next stage of the research, namely pilot implementation and empirical testing. The final validated model was therefore used as the basis for evaluating usability, teacher and student acceptance, scoring consistency, implementation workload, and the effectiveness of the AI-resilient assessment approach in measuring competency within a Generative AI-enabled educational environment.

3.5 Practicality Results

The results of the practicality evaluation indicate that the implementation of the proposed model needs to be considered not only from the perspective of assessment quality but also from the perspective of usability and workload. The model requires teachers to evaluate several forms of evidence, including authentic tasks, process documentation, AI-use disclosure, final products or performances, oral defense, and student reflection. Although this approach provides richer evidence of competency, it can potentially require more preparation and evaluation time than conventional assessment. Therefore, the practicality assessment examined whether the additional assessment components could be managed within the normal responsibilities of teachers.

Regarding teacher implementation time, the evaluation focused on the amount of time required to prepare assessment tasks, explain the assessment procedures, review process evidence, evaluate final products, conduct oral defense, and complete competency judgments. The findings indicate that the model introduces additional assessment activities compared with conventional assessment, particularly because teachers must evaluate evidence collected throughout the learning process. However, the structured sequence and standardized assessment procedures can reduce unnecessary repetition when the model is implemented consistently. The use of predefined rubrics and assessment templates can also help teachers organize evidence and make competency judgments more efficiently. The results suggest that the model is feasible when assessment components are appropriately integrated into existing learning activities rather than implemented as completely separate activities.

The ease of using the assessment rubric was another important aspect of practicality(Allen & Tanner, 2006). The rubric was designed to translate the eight dimensions of the model authenticity, process evidence, higher-order thinking, AI transparency, AI literacy, oral defense, reflection, and competency demonstration into observable assessment criteria. Teachers were asked to evaluate whether the indicators were sufficiently clear, whether performance levels could be distinguished, and whether the rubric could support consistent scoring. The practicality results suggest that a structured rubric can assist teachers in evaluating multiple sources of evidence and reduce ambiguity when determining competency achievement. Nevertheless, the number of indicators needs to be controlled to prevent the rubric from becoming unnecessarily complicated. Clear descriptors and examples of expected student performance are therefore important for ensuring that the rubric remains practical.

The student understanding component examined whether students could understand the purpose, procedures, assessment criteria, AI-use requirements, and evidence that they were expected to provide. This aspect is particularly important because the model introduces several requirements that may be unfamiliar to students, including process documentation, AI-use disclosure, oral defense, and reflection. The evaluation indicates that students need clear instructions before beginning the assessment. Students should understand that AI transparency does not necessarily mean that AI use is automatically prohibited; rather, they are expected to disclose and explain their use of AI according to the assessment requirements. Clear examples, assessment guidelines, and rubric explanations can help students understand the expectations and reduce uncertainty during implementation.

The evaluation of administrative feasibility considered the extent to which the model could be incorporated into existing curriculum and assessment procedures. The model requires teachers to manage several forms of evidence, which may increase the need for systematic documentation and record keeping. To maintain feasibility, evidence collection should be integrated into existing learning management systems, portfolios, assignment submission procedures, or other institutional assessment platforms where available. The assessment model should also avoid requiring teachers to create completely new administrative procedures for every assessment. Standardized templates for AI-use disclosure, process documentation, reflection, and competency judgment can simplify administration and improve consistency across classes or courses.

Another important aspect concerns technology requirements. The model is designed to remain functional without depending entirely on sophisticated AI-detection systems or specialized technological infrastructure. Basic digital tools may be sufficient for collecting process documentation, AI-use disclosures, student products, reflection records, and other assessment evidence. Where institutional infrastructure permits, learning management systems, digital portfolios, document histories, or electronic rubrics can be used to facilitate evidence collection and evaluation. However, the model should remain adaptable to different levels of technological availability so that institutions with limited technological resources can still implement its core principles.

3.6 Interpretation of Practicality Findings

The practicality findings indicate that the AI-Resilient Competency Assessment Model has the potential to be implemented in authentic educational settings, provided that the assessment procedures are appropriately organized. The findings are particularly important because the value of an assessment model is determined not only by its theoretical validity but also by whether teachers and students can use it consistently without creating an unreasonable burden. The practicality evaluation therefore provides evidence concerning the operational feasibility of the model and helps determine whether the proposed assessment framework can move beyond theoretical development toward sustainable educational practice.

The findings regarding teacher implementation time suggest that the AI-resilient model may require more assessment effort than conventional assessment because it collects multiple forms of evidence. Teachers are required to consider not only the final student product but also the learning process, AI-use disclosure, performance, oral explanation, and reflection(Zimmerman, 2018). This additional evidence collection can initially increase preparation, monitoring, and scoring time. However, the increased workload should be interpreted in relation to the quality of assessment evidence obtained. The purpose of the model is not to make assessment longer for its own sake, but to replace a potentially unreliable single-source assessment with a more comprehensive evidence-based process. When process documentation, oral defense, and reflection are incorporated into regular learning activities, the additional workload can be reduced because these components become integrated parts of the learning process rather than separate assessment activities.

The findings concerning rubric usability indicate that a well-structured rubric is essential for making the model practical. Because the model contains eight dimensions, namely authenticity, process evidence, higher-order thinking, AI transparency, AI literacy, oral defense, reflection, and competency demonstration, teachers need clear criteria to distinguish different levels of student performance. Without clear descriptors, the richness of the model could increase subjective judgment and scoring inconsistency. Therefore, the practicality findings support the use of concise, observable, and competency-oriented indicators. The rubric should function as a decision-support instrument for teachers rather than as an additional administrative burden. This interpretation also suggests that future implementation should provide examples of student performance for different competency levels so that teachers can apply the rubric more consistently.

The findings related to student understanding demonstrate that the success of the model depends on students clearly understanding what is being assessed and why each type of evidence is required. The introduction of AI-use disclosure, process evidence, oral defense, and reflection represents a change from conventional assessment practices in which students may only be required to submit a final answer. Students therefore need explicit instructions concerning the acceptable use of AI, the information that must be disclosed, the types of process evidence that should be retained, and the criteria used to evaluate their competency. This is particularly important because AI transparency should not be interpreted as a mechanism for automatically penalizing students for using AI. Instead, it should encourage students to use AI responsibly, critically evaluate its outputs, and demonstrate ownership of their learning.

The findings on administrative feasibility suggest that the model is more sustainable when its components are integrated with existing assessment and learning systems. If teachers are required to maintain separate records for every assessment component, the administrative burden may become excessive. Therefore, the model should encourage the use of integrated documentation, standardized forms, digital portfolios, learning management systems, or electronic rubrics where available. This approach can allow process evidence, AI-use disclosure, reflection, and final assessment results to be stored systematically. Administrative feasibility is consequently not merely a technical issue but an important factor affecting whether the model can be adopted consistently at the institutional level.

The findings concerning technology requirements further indicate that AI-resilient assessment should not depend entirely on sophisticated AI-detection technologies. This is a significant implication of the study because the central purpose of the model is to maintain assessment validity even when students have access to Generative AI. The model can use relatively accessible technologies for collecting documentation, submitting assignments, recording AI use, managing rubrics, and storing assessment evidence. Institutions with more advanced technological infrastructure may integrate learning management systems, digital portfolios, version histories, or other automated tools, whereas institutions with limited resources can implement the core assessment principles using simpler digital or even partially manual procedures.

Taken together, the findings suggest that the practicality of the model depends on achieving an appropriate balance between assessment comprehensiveness and implementation efficiency(de Olde et al., 2018). The model deliberately collects multiple forms of evidence because a final product alone may not adequately demonstrate authentic competency in the GenAI era. However, requiring every dimension to be implemented at maximum intensity for every assignment could create unnecessary workload. The model should therefore be applied flexibly according to the learning outcomes and characteristics of each assessment. For example, a practical course may emphasize performance demonstration and oral defense, while a project-based course may place greater emphasis on process evidence, higher-order thinking, and reflection.

An important interpretation is that the AI-resilient model represents a redesign of assessment rather than an addition of assessment tasks. This distinction is essential for addressing concerns about teacher workload. If oral defense, reflection, and process documentation are simply added to existing essays, projects, examinations, and assignments, the model could become impractical. However, if these components are incorporated into the existing assessment workflow, they can replace weaker forms of evidence and provide richer information without proportionally increasing the total workload. For example, a project can simultaneously generate the final product, process evidence, AI-use disclosure, reflection, and competency evidence rather than requiring five separate assignments.

The findings also have implications for teacher professional development. The successful implementation of the model requires teachers to understand not only how to use the rubric but also how to design authentic tasks, interpret process evidence, evaluate AI use, conduct efficient oral defense, and integrate multiple sources of evidence into a final competency judgment. Therefore, teacher training should be considered an important implementation strategy. Professional development can help teachers develop assessment literacy and AI literacy while reducing uncertainty regarding the practical application of the model.

4. CONCLUSION

This study successfully developed an AI-Resilient Competency Assessment Model designed to maintain the validity and authenticity of competency-based assessment in an educational environment where Generative Artificial Intelligence is increasingly accessible. The model responds to the limitations of conventional assessment, particularly the dependence on final products that may not adequately represent students' genuine competencies when AI-assisted outputs are used. The final model integrates several interconnected components, including authentic and contextual tasks, process-based evidence, higher-order thinking, AI-use transparency, AI literacy, product and performance assessment, oral defense, and student reflection. These components provide multiple sources of evidence for determining whether students have achieved the competencies specified in the curriculum. The model therefore shifts assessment from evaluating only what students produce toward evaluating how they think, perform, explain, reflect, and responsibly use AI. Expert validation, pilot implementation, and empirical testing provide the basis for determining the model's validity, reliability, practicality, and effectiveness; however, these claims should be reported according to the actual statistical findings obtained in the study. The model has practical implications for schools, universities, vocational institutions, and teacher professional development by supporting a transition from product-centered assessment toward evidence-rich competency assessment. Nevertheless, the study may be limited by its institutional context, sample size, implementation period, differences in teachers' AI literacy, and the rapidly changing nature of GenAI technologies. Future research should test and refine the model across different educational levels, disciplines, institutions, countries, AI environments, and longitudinal settings.

REFERENCES

Agnell, F. (2015). *AI adoption in the education system International insights and policy considerations for Italy*. OECD.

Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE—Life Sciences Education*, 5(3), 197–203.

Banta, T. W., & Palomba, C. A. (2014). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. John Wiley & Sons.

Bednar, A. K., Cunningham, D., Duffy, T. M., & Perry, J. D. (2013). Theory into practice: How do we link? In *Constructivism and the technology of instruction* (pp. 17–34). Routledge.

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: a conceptual analysis and model. *Journal of Applied Psychology*, *91*(6), 1189.

Carver, J., Jaccheri, L., Morasca, S., & Shull, F. (2004). Issues in using students in empirical studies in software engineering education. *Proceedings. 5th International Workshop on Enterprise Networking and Computing in Healthcare Industry (IEEE Cat. No. 03EX717)*, 239–249.

Curry, L., & Docherty, M. (2017). Implementing competency-based education. *Collected Essays on Learning and Teaching*, *10*, 61–73.

de Olde, E. M., Sautier, M., & Whitehead, J. (2018). Comprehensiveness or implementation: Challenges in translating farm-level sustainability assessments into action for sustainable development. *Ecological Indicators*, *85*, 1107–1112.

Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical Assessment, Research & Evaluation*, *17*(2), n2.

Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, *1*, 3–31.

Haudek, K. C., Kaplan, J. J., Knight, J., Long, T., Merrill, J., Munn, A., Nehm, R., Smith, M., & Urban-Lurain, M. (2011). Harnessing technology to improve formative assessment of student conceptions in STEM: forging a national network. *CBE—Life Sciences Education*, *10*(2), 149–155.

Havnes, A., Smith, K., Dysthe, O., & Ludvigsen, K. (2012). Formative assessment and feedback: Making learning visible. *Studies in Educational Evaluation*, *38*(1), 21–27.

Kay, S. (2004). The Move from Oral Evidence to Written Evidence:'The Law Is Always Too Short and Too Tight for Growing Humankind'. *J. Int'l Crim. Just.*, *2*, 495.

Kouwenhoven, W. (2009). Competence-based curriculum development in higher education: A globalised concept? *Technology Education and Development*, *8*(2), 1–22.

Kumar, S., & Rani, R. U. (n.d.). Artificial. *TOWARDS OPEN, DECENTRALISED AI ECOSYSTEMS*, 41.

Leroy, J. L., Ruel, M., Frongillo, E. A., Harris, J., & Ballard, T. J. (2015). Measuring the food access dimension of food security: a critical review and mapping of indicators. *Food and Nutrition Bulletin*, *36*(2), 167–195.

Leslie, D. (2019). Understanding artificial intelligence ethics and safety. *ArXiv Preprint ArXiv:1906.05684*.

Mendoza, N. B., Cheng, E. C. K., & Yan, Z. (2022). Assessing teachers' collaborative lesson planning practices: Instrument development and validation using the SECI knowledge-creation model. *Studies in Educational Evaluation*, *73*, 101139.

Molloy, E., & Boud, D. (2012). Changing conceptions of feedback. In *Feedback in higher and professional education* (pp. 11–33). Routledge.
Pareto, L., Haake, M., Lindström, P., Sjödén, B., & Gulz, A. (2012). A teachable-agent-based game affording collaboration and competition: Evaluating math comprehension and motivation. *Educational Technology Research and Development*, *60*(5), 723–751.
Pedro, F., Subosa, M., Rivas, A., & Valverde, P. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development*.
Rikhardsson, P. M. (2017). Information systems for corporate environmental management accounting and performance measurement. In *Sustainable Measures* (pp. 132–150). Routledge.
Samwel Mwasalwiba, E. (2010). Entrepreneurship education: a review of its objectives, teaching methods, and impact indicators. *Education+ Training*, *52*(1), 20–47.
Schneider, M. C., & Gowan, P. (2013). Investigating teachers' skills in interpreting evidence of student learning. *Applied Measurement in Education*, *26*(3), 191–204.
Schwartz, D. L., Lin, X., Brophy, S., & Bransford, J. D. (2013). Toward the development of flexibly adaptive instructional designs. In *Instructional-design theories and models* (pp. 183–213). Routledge.
Steiner, G., & Posch, A. (2006). Higher education for sustainability by means of transdisciplinary case studies: an innovative approach for solving complex, real-world problems. *Journal of Cleaner Production*, *14*(9–11), 877–890.
Thompson, J. E. (2021). *Influencing Trust in Human and Artificial Intelligence Teaming Through Heuristics*.
Tzirides, A. O. O., Cope, B., & Kalantzis, M. (2020). Contextual AI meets generative AI: Addressing the literacy crisis. In *Artificial Intelligence in Literacy Education* (pp. 3–25). Routledge.
Zimmerman, M. (2018). *Teaching AI: Exploring new frontiers for learning*. ASCD.