Online-Journals.org

iJEP International Journal of Engineering Pedagogy

iJEP | eISSN: 2192-4880 | Vol. 16 No. 1 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijep.v16i1.60479

PAPER

Designing Formative Adaptive Assessment for Engineering Education: Integrating Computerized Adaptive Testing and Competency-Based Diagnostic Modelling

Mohamed El Msayer[1] (✉),
Bouchra Bouihi[1],
Abdelmajid Bousselham[1],
Essaadia Aoula[1],
Adel Deraoui[2]

[1]Laboratory 2IACS ENSET Mohammedia, Hassan II University of Casablanca, Casablanca, Morocco

[2]Regional Center for Education and Training, Casablanca, Morocco

mohamed.elmsayer-etu@etu.univh2c.ma

ABSTRACT

Assessing competencies in engineering education increasingly requires digital assessment approaches that support learning regulation, instructional decision-making, and educational quality, rather than focusing solely on measurement efficiency. Computerized adaptive testing (CAT), grounded in Item Response Theory (IRT), provides a robust methodological foundation for personalized assessment. However, its pedagogical effectiveness in formative contexts depends critically on curriculum alignment, diagnostic capacity, and adaptive control strategies. This study proposes and evaluates a formative adaptive assessment framework for engineering education that integrates an IRT-based CAT engine with a Bayesian network–based diagnostic component. The framework is designed to support competency-oriented feedback, learning monitoring, and instructional interpretation within a curriculum-aligned assessment structure. Assessment relies on dichotomous multiple-choice items explicitly aligned with engineering learning outcomes, while item selection dynamically adapts to learners' evolving proficiency estimates. In parallel, probabilistic diagnostic modelling prioritizes under-assessed competencies throughout the adaptive process. Item calibration was conducted using empirical data collected from 612 university students in computer science, and system performance was examined through a simulation-based evaluation involving 500 simulated learners. Results demonstrate high estimation accuracy (r = 0.912) and satisfactory reliability for formative use across most learner profiles. Reduced precision at the extremes of the proficiency continuum and imbalances in item exposure were also observed, highlighting structural limitations primarily related to item bank coverage and curriculum representation rather than to the adaptive algorithms themselves. Overall, the proposed framework positions adaptive assessment as a pedagogically grounded tool for formative learning support, instructional decision-making, and quality assurance in engineering education.

Msayer, M. E., Bouihi, B., Bousselham, A., Aoula, E., Deraoui, A. (2026). Designing Formative Adaptive Assessment for Engineering Education: Integrating Computerized Adaptive Testing and Competency-Based Diagnostic Modelling. *International Journal of Engineering Pedagogy (iJEP)*, 16(1), pp. 133–154. https://doi.org/10.3991/ijep.v16i1.60479

Article submitted 2026-01-05. Revision uploaded 2026-01-16. Final acceptance 2026-01-16.

KEYWORDS
engineering education, formative adaptive assessment, computerized adaptive testing (CAT), Item Response Theory (IRT), curriculum alignment

1 INTRODUCTION

Assessment plays a central role in the quality of engineering education, particularly in higher education, where it directly informs pedagogical decision-making, student orientation, and the certification of competencies. In engineering programs, assessment is not merely a mechanism for measuring achievement, it also serves as a key lever for learning regulation, curriculum alignment, and quality assurance. With the widespread adoption of digital learning environments, assessment practices have progressively evolved toward more flexible, interactive, and learner-centered approaches that better accommodate heterogeneous learner profiles and complex competency structures [1], [2]. Within this context, artificial intelligence (AI) offers valuable opportunities to design assessment systems capable of adapting to individual learners and providing accurate, actionable diagnostic information to support both teaching and learning processes [3].

Traditional assessment approaches in engineering education are largely based on fixed-form tests that present identical sets of items to all learners. While such tests are relatively easy to administer, they exhibit important limitations in educational contexts characterized by substantial variability in learners' prior knowledge and competency levels. Fixed tests frequently include items that are either too easy or too difficult for a given learner, thereby reducing measurement efficiency, learner engagement, and the pedagogical value of assessment outcomes [4]. To address these limitations, computerized adaptive testing (CAT) has emerged as a robust alternative, enabling the dynamic selection of items based on an examinee's estimated proficiency level and reducing test length while maintaining high measurement precision [5]. From an assessment for learning perspective, CAT is particularly well suited to formative assessment, as it supports more individualized and efficient monitoring of learner progress [6].

From an educational assessment design perspective, Item Response Theory (IRT) provides the methodological foundation of CAT systems. IRT models, and in particular the three-parameter logistic (3PL) model, enable the estimation of a latent competency representing learners' performance levels based on responses to dichotomous multiple-choice items, while accounting for item difficulty, discrimination, and pseudo-guessing parameters. These models have demonstrated their effectiveness for adaptive competency assessment across a wide range of educational and professional contexts [7]. However, CAT systems relying exclusively on IRT typically yield a single global ability estimate (θ), offering limited insight into the specific components or sub-skills underlying learner performance. This limitation constrains their pedagogical usefulness in formative assessment contexts, where fine-grained diagnostic information is essential for supporting learning regulation and instructional adjustment [5]. To enhance the adaptive and diagnostic capacity of assessment systems, recent research has explored the integration of AI techniques such as neural networks, reinforcement learning, and Bayesian networks within CAT frameworks [5], [6], [8]. These approaches aim to improve estimation accuracy, diversify item administration, and strengthen item bank security. In particular, Bayesian networks provide a powerful probabilistic framework for modelling relationships among competencies, incorporating response histories, and dynamically updating learners' knowledge states [7]. Despite this potential, much of the existing literature employs Bayesian networks primarily in a descriptive or auxiliary role, without a clear and operational integration with the psychometric mechanisms governing adaptive item selection and learning-oriented feedback. Despite significant advances in adaptive assessment research [9], [10], several gaps remain, especially with respect to formative assessment in engineering education. Most existing CAT systems have been developed primarily for summative purposes and focus on producing a global proficiency score, offering limited diagnostic information to support learning processes and instructional decision-making [5]. Approaches that integrate AI and Bayesian networks in particular often lack a coherent methodological articulation with IRT models, especially regarding the explicit role of probabilistic diagnostics in adaptive item selection strategies [11]. Furthermore, relatively few studies adopt a joint validation approach that combines empirical data for item calibration with controlled simulations to examine robustness, estimation stability, and item bank behavior. Critical issues related to fairness, coverage of the full ability continuum, and item exposure management therefore remain insufficiently addressed in large-scale formative adaptive assessment systems [5]. These limitations highlight the need for adaptive assessment systems in engineering education that move beyond the sole production of global proficiency scores and explicitly serve formative pedagogical purposes. In digital engineering education contexts, effective adaptive assessment must support diagnostic interpretability, learning personalization, and responsible measurement control while remaining closely aligned with instructional objectives and curriculum structures.

In response to these challenges, this study proposes a formative adaptive assessment framework for engineering education that integrates three complementary components. First, IRT-3PL is used to support curriculum-aligned item calibration and adaptive item selection, ensuring measurement rigor and coherence with learning outcomes. Second, a competency-based diagnostic component, implemented through a Bayesian network, enables dynamic updating of learners' competency profiles, models relationships among knowledge components, and generates individualized feedback that is pedagogically interpretable. Third, a hybrid evaluation strategy combines empirical data for item calibration and feasibility analysis with simulation-based evaluation of system behavior, estimation reliability, and item bank sustainability. Methodologically, the study adopts an integrated approach that brings together principles of assessment design, pedagogical engineering, and educational data modelling to address formative assessment challenges in higher education. Beyond supporting individual learner feedback, the diagnostic outputs generated by the framework are explicitly designed to inform instructional decision-making at both course and curriculum levels. By enabling instructors to identify systematic competency gaps, prioritize remediation strategies, and adapt instructional sequencing, the framework positions adaptive assessment as a pedagogical decision-support tool rather than a purely measurement-oriented system.

The remainder of this paper is organized as follows. Section 2 presents the theoretical background framing formative adaptive assessment within engineering education. Section 3 describes the design and methodological implementation of the proposed framework, including its pedagogical rationale, adaptive assessment mechanisms, and diagnostic components. Section 4 reports the results of the empirical calibration and simulation-based evaluation. Finally, Section 5 discusses the findings from an engineering education perspective, focusing on implications for formative assessment, learning regulation, and instructional decision-making, and outlines directions for future research.

2 BACKGROUND

2.1 CAT in engineering education

Computerized adaptive testing emerged in the 1970s and has evolved steadily alongside advances in computing technologies and modern measurement models [3]. CAT is based on an iterative assessment process in which test items are dynamically selected according to a learner's estimated proficiency level, with the dual objective of maximizing measurement precision while minimizing test length [5]. In the context of engineering education, CAT has gained increasing attention as a means of supporting efficient, learner-sensitive assessment in learning environments characterized by heterogeneous competency profiles and complex curricular outcomes. Table 1 summarizes the main methodological stages reported in prior research on IRT-based CAT systems, covering the full development cycle from design considerations to operational implementation [12]. Beyond their technical formulation, these stages reflect key pedagogical, curricular, and organizational decisions, including the definition of assessment purposes, construct specification, curriculum alignment, and institutional readiness.

Table 1. Methodological stages and core components of IRT-based CAT systems

Methodological Stage	Core Component	Description
Feasibility and pedagogical design rationale	Assessment purpose and CAT relevance	Prior studies justify the use of CAT in relation to formative assessment goals, targeted engineering competencies, learner characteristics, and operational constraints such as test length, feedback needs, and reporting precision.
	Institutional and instructional readiness	The literature emphasizes the importance of assessment expertise, secure digital delivery infrastructures, and sustainable item bank governance to support pedagogically meaningful and scalable CAT implementation.
Curriculum-aligned item bank development and calibration	Item bank construction	Research consistently reports systematic item development guided by curriculum specifications, expert validation, and learning outcomes to ensure construct coverage, fairness, and alignment with engineering competency frameworks.
	IRT-based calibration	Empirical calibration using appropriate unidimensional or multidimensional IRT models (e.g., 1PL, 2PL, 3PL, MIRT) is identified as a prerequisite for reliable and information-driven adaptive assessment.
Measurement and assessment quality assurance	Model assumptions	Studies routinely examine dimensionality, local independence, and monotonicity to ensure the validity, interpretability, and educational meaningfulness of proficiency estimates (θ).
	Model–data fit	Item- and person-fit analyses are conducted to identify misfit, refine the item bank, and support stable and equitable adaptive assessment behavior across learner populations.

(Continued)

Table 1. Methodological stages and core components of IRT-based CAT systems *(Continued)*

Methodological Stage	Core Component	Description
Adaptive assessment and decision framework	Item selection strategies	Prior CAT implementations predominantly rely on information-based item selection rules (e.g., maximum Fisher information, Kullback–Leibler divergence), often combined with content balancing or exposure control to support diagnostic relevance and fairness.
	Ability estimation procedures	Iterative proficiency estimation using maximum likelihood and/or Bayesian estimators (EAP, MAP) is standard practice, supporting reliable learning monitoring and instructional interpretation.
	Termination rules	Fixed-length, precision-based (SE(θ)), or hybrid stopping criteria are documented as mechanisms to balance measurement efficiency, reliability, and formative usability.

A typical CAT system comprises five core components: a starting point, a calibrated item bank, an item selection algorithm, a scoring mechanism, and a termination criterion, as illustrated in Figure 1. The item bank constitutes the curricular content of the assessment, while the remaining components correspond to adaptive control mechanisms that govern item selection, proficiency estimation, and test termination [5]. Following the establishment of an initial estimate, the adaptive cycle iteratively selects items, updates proficiency estimates, and evaluates stopping criteria until the assessment process is completed.

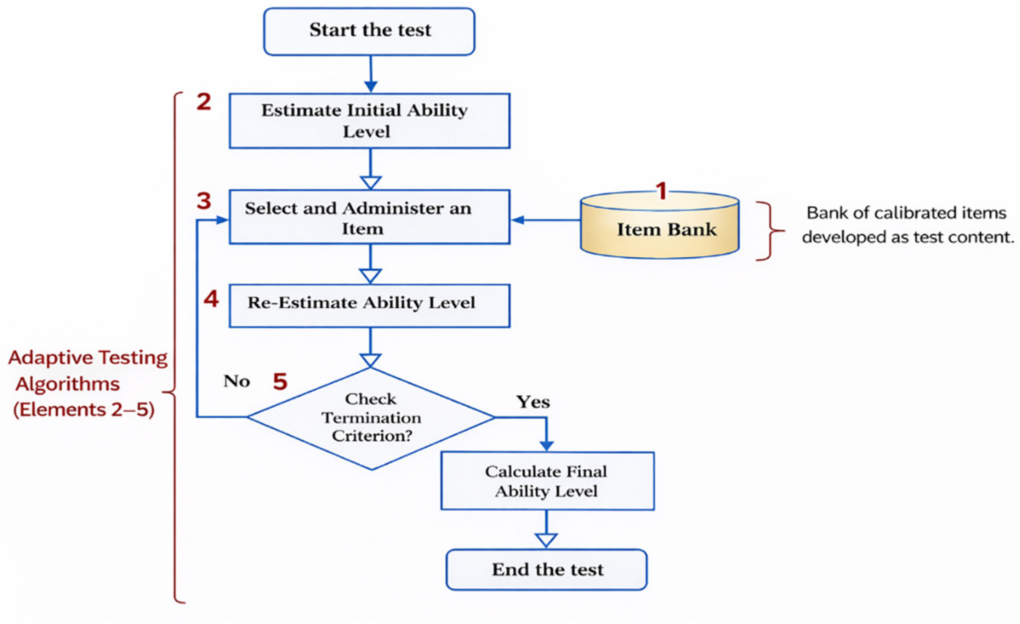

Fig. 1. Overview of the CAT process

Item banks are traditionally developed by domain experts; however, recent advances in natural language processing and automatic item generation have enabled the scalable production of items with psychometric properties comparable to manually authored questions [13]. The development of an item bank requires rigorous evaluation procedures, including the examination of dimensionality, local independence, model data fit, and the stability of item functioning across learner subpopulations [14]. These requirements are particularly critical in formative assessment contexts, where equitable measurement across heterogeneous proficiency levels is essential. Insufficient item bank coverage may result in reduced precision, especially at the lower and upper extremes of the ability continuum. Accordingly, the literature commonly recommends item bank sizes ranging from five to ten times the number of items administered during adaptive testing [7].

2.2 IRT as a foundation for adaptive assessment

The methodological foundation of CAT systems is IRT, also referred to as latent trait theory, which models the probability of a correct response as a function of both learner characteristics and item parameters [15]. This relationship is commonly expressed as:

$$P(i = a \mid \theta, \gamma) = F(\theta, \gamma)$$

where θ denotes the learner's latent proficiency and γ represents item parameters such as difficulty, discrimination, and pseudo-guessing.

For dichotomous items, particularly multiple-choice questions commonly used in engineering education, IRT models assign binary scores to responses. The Rasch model (1PL) represents the simplest formulation, while the 2PL and 3PL models extend this framework by incorporating discrimination and pseudo-guessing parameters, respectively. The 3PL model is especially appropriate for multiple-choice assessments, provided that sufficient empirical data are available to ensure stable parameter estimation [16]. For this reason, the present study adopts the 3PL model as the basis for adaptive item calibration and proficiency estimation. IRT also includes models for polytomous response formats, such as Likert-scale items, including the graded response model, the partial credit model, and the nominal response model [17].

2.3 AI in adaptive assessment systems

Recent research has highlighted the potential of AI to enhance both the accuracy and the pedagogical relevance of adaptive assessment systems. Reinforcement learning approaches enable dynamic adjustment of item difficulty based on observed learner performance [18], while neural network architectures, including feed-forward models [19], long short-term memory (LSTM) networks [20], and Bayesian probabilistic models [10], support more refined learner profiling through the analysis of response histories. Despite these advances, the majority of operational adaptive assessment systems continue to rely on IRT as a robust and interpretable framework for estimating learner proficiency. Classical item selection strategies, such as Fisher information maximization and Kullback–Leibler divergence–based approaches, remain widely used to control measurement error and ensure estimation accuracy [9]. Within engineering education, this combination of measurement rigor in educational assessment and adaptive flexibility is particularly important for supporting meaningful formative assessment practices.

Item selection approaches. The literature reports a wide range of item selection approaches in adaptive assessment systems, including IRT-based strategies, neural networks, Bayesian networks, decision trees, and genetic algorithms [8], [21], [22]. IRT-based methods remain central due to their strong theoretical foundations and interpretability, while AI-based approaches offer increased flexibility for modelling complex relationships among items, competencies, and learning trajectories. Together, these hybrid approaches reflect a broader evolution toward intelligent adaptive assessment systems that seek to balance measurement precision, diagnostic relevance, and operational efficiency [23]. In the context of engineering education, such systems are increasingly conceptualized not merely as measurement tools but as integral components of pedagogical engineering, supporting learning regulation, instructional decision-making, and curriculum alignment.

3 METHODOLOGY

3.1 Design of a formative adaptive assessment framework for engineering education

The overall architecture of the proposed system, illustrated in Figure 2, is conceived as a formative adaptive assessment framework explicitly oriented toward engineering education. Rather than focusing solely on measurement efficiency, the framework is designed to support learning regulation, competency diagnosis, and instructional decision-making through an integrated and iterative assessment process. It is structured around five interconnected functional components that collectively enable adaptive assessment, pedagogically meaningful diagnosis, and actionable feedback within a coherent formative assessment loop.

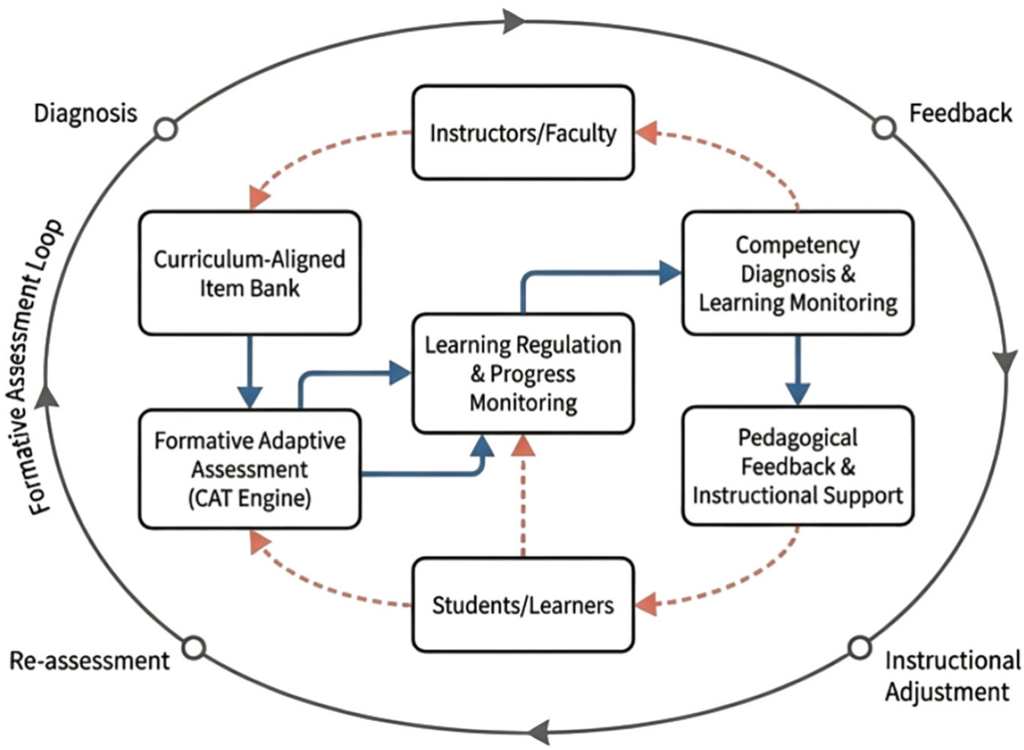

Fig. 2. Architecture of a formative adaptive assessment framework for engineering education

The first component, the Curriculum-Aligned Item Bank (01), constitutes the curricular and pedagogical foundation of the system. It enables instructors to design items explicitly aligned with engineering learning outcomes and disciplinary competency frameworks. Each item is organized within discipline-specific item banks. Prior to operational use, items undergo calibration to ensure measurement quality, parameter stability, and fairness across heterogeneous learner populations. In pedagogical terms, this component ensures that adaptive assessment remains firmly anchored in curriculum intentions and learning objectives. At the core of the architecture is the Formative Adaptive Assessment component (02), implemented through a CAT engine. This component manages adaptive item selection, test administration, response scoring, and termination decisions. By continuously updating learners' proficiency estimates based on their responses, the CAT engine dynamically selects items that are most informative for each learner. From an educational perspective, this adaptivity supports formative assessment by limiting unnecessary testing burden while maintaining sufficient precision to inform instructional and learning decisions. The Learning Regulation and Progress Monitoring component (03) embodies the formative use of assessment results. It supports the repeated and iterative administration of adaptive assessments across different stages of the learning trajectory. As reflected by the feedback loops shown in Figure 2, assessment outcomes are systematically reintegrated into the learning process. This design enables continuous monitoring of learner progress and supports timely pedagogical adjustments, positioning assessment as a tool for learning regulation rather than as an exclusively summative mechanism. The Competency Diagnosis and Learning Monitoring component (04) provides fine-grained diagnostic information through the integration of a Bayesian network. This component models probabilistic relationships among targeted engineering competencies and continuously updates learners' competency states based on response histories generated by the adaptive assessment engine. Beyond a single global proficiency estimate, this diagnostic process supports pedagogical interpretation of learner performance at the level of underlying competencies, thereby enhancing the formative value of assessment information. Finally, the Pedagogical Feedback and Instructional Support component (05) synthesizes information derived from adaptive assessment and diagnostic monitoring to generate individualized feedback. This feedback is directed both to students/learners, supporting self-regulation and informed learning progression, and to instructors/faculty, informing instructional decision-making, curriculum adjustment, and quality assurance practices within engineering programs.

As depicted in Figure 2, the proposed system operates as a closed-loop formative assessment framework that integrates automated adaptive assessment and diagnostic processes with human pedagogical judgment. Solid arrows represent automated measurement and AI-supported processes, whereas dashed arrows indicate human decision-making by students and instructors. Together, these interactions position adaptive assessment as a pedagogical support system that facilitates continuous formative assessment, learning regulation, and evidence-based instructional practice in engineering education.

3.2 Data collection and sample characteristics

Empirical data were collected in the context of computer science assessments within an engineering education setting, specifically through entrance examinations administered to candidates applying for a master's program at ENSET Mohammedia. These assessments were designed to evaluate core disciplinary competencies relevant to studies in computer science. Each examination comprised 40 dichotomous multiple-choice items, and the overall development process resulted in a curriculum-aligned item bank of 200 items. This item bank was constructed to represent a range of content areas and competency levels aligned with the target curriculum, thereby providing the empirical basis required for reliable calibration and subsequent adaptive use.

The empirical sample consisted of 612 students with heterogeneous academic backgrounds in computer science and related disciplines. Participants were drawn from several Moroccan and international universities and were selected in accordance with the program's established admission criteria. This diversity reflects the variability typically encountered in engineering education contexts and supports the evaluation of measurement quality and fairness across learner subpopulations. Learners' responses were collected using standardized answer sheets and processed through an automated optical scanning system. These empirical response data were used exclusively for item calibration, feasibility analysis, and verification of measurement assumptions underpinning the adaptive assessment system.

3.3 Development and calibration of the curriculum-aligned item bank

The final item bank consisted of 200 dichotomous multiple-choice items, calibrated using the three-parameter logistic IRT-3PL model. This calibration process provided the measurement parameters required to support adaptive item selection and ability estimation within the CAT framework. Figure 3 presents the distributions of item discrimination, difficulty, and pseudo-guessing parameters obtained from this calibration. These parameters formed the measurement foundation of the adaptive assessment system and were jointly used to guide adaptive item selection and compute learners' proficiency estimates [14].

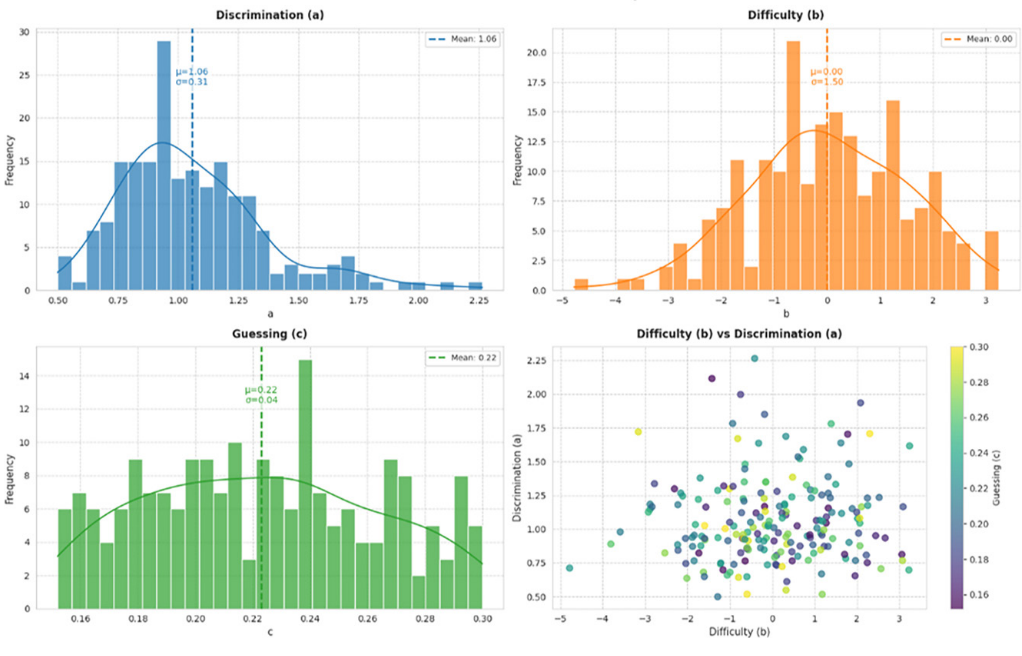

Fig. 3. Distribution of discrimination, difficulty, and guessing parameters in the item bank

The calibration process also enabled computation of item information functions and the overall test information function, which quantify measurement precision across the ability continuum. These information functions play a central role in adaptive assessment, as they inform item selection decisions and allow evaluation of expected measurement error at different proficiency levels [14]. Figure 4 illustrates the test information function and the associated standard error of measurement across the latent ability continuum, as obtained from calibration of the item bank using the IRT-3PL model. During adaptive administration, item selection was guided by the information provided by each item at the learner's current ability estimate, ensuring that assessment focused on the most informative items for each individual. Test administration was terminated once a predefined precision threshold was reached, allowing reliable ability estimation while limiting test length. Overall, the distribution of test information indicates that the calibrated item bank provides adequate coverage of the targeted competency domain and supports the effective implementation of a formative computerized adaptive assessment system.

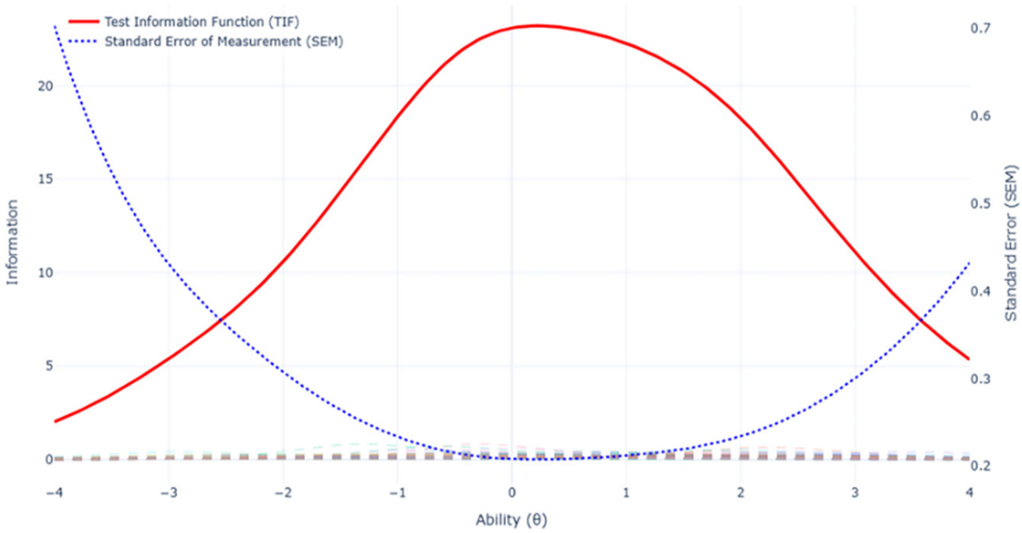

Fig. 4. Test information function and standard error of measurement across the ability continuum

3.4 Adaptive item selection and diagnostic integration

Adaptive item selection was primarily governed by a Fisher information maximization criterion, with the objective of minimizing measurement error around the learner's current latent ability estimate θ. At each stage of the adaptive process, the item providing the highest information at the estimated ability level was selected, thereby supporting efficient convergence toward stable proficiency estimates. In parallel, a Bayesian network was integrated as a complementary diagnostic mechanism [10]. This network modelled probabilistic relationships among targeted competencies and incorporated learners' response histories to update posterior competency states in real time. Posterior probabilities inferred from the Bayesian network were used to (i) prioritize under-assessed competencies, (ii) guide item selection toward diagnostically relevant content areas, and (iii) support the generation of individualized formative feedback [10]. This hybrid integration preserves the measurement rigor of IRT-based adaptive selection while enriching the assessment process with diagnostically meaningful constraints, ensuring coherence between measurement precision and formative objectives [5].

3.5 Ability estimation strategy in the adaptive assessment process

Ability estimation was conducted within the adaptive testing process using the IRT-3PL framework adopted for item calibration. Rather than reiterating model properties, this section focuses on the operational estimation strategy implemented to ensure stable and accurate proficiency estimates throughout the adaptive assessment. A two-stage hybrid estimation approach was employed. During the initial stages of the test, when response information was limited, Expected A Posteriori (EAP) estimation with a standard normal prior distribution $N(0,1)$ was used to stabilize early estimates and prevent extreme values. As additional responses were collected and item information increased, the estimation procedure transitioned to Maximum Likelihood Estimation (MLE), allowing estimates to rely more directly on observed response patterns and improving precision in later test stages. This strategy is consistent with established practices in adaptive testing and supports reliable estimation across different phases of assessment [24].

3.6 Simulation of the adaptive assessment system

A simulation study was conducted to examine the robustness and formative usefulness of the proposed adaptive assessment framework under controlled conditions, in line with common validation practices in engineering education research. A simulated cohort of 500 learners was generated, with proficiency levels drawn from a standard normal distribution ($\theta \sim N(0,1)$), reflecting the heterogeneity typically observed in engineering student populations. Each adaptive assessment began with a neutral initial proficiency estimate ($\theta = 0$). During the assessment process, items were selected iteratively based on their instructional and measurement relevance at the learner's current proficiency level. Learner responses were simulated using the IRT-3PL model, and proficiency estimates were updated at each step using a Maximum A Posteriori (MAP) approach informed by the Bayesian diagnostic component. This integration ensured that adaptive decisions were guided not only by measurement considerations but also by competency-level diagnostic information. The adaptive process was terminated either when a predefined level of precision suitable for formative interpretation was reached (standard error ≤ 0.30) or when a maximum of 30 items had been administered. From a pedagogical standpoint, these stopping rules reflect a balance between assessment efficiency and the need for sufficiently stable feedback to support learning regulation and instructional decision-making in engineering education contexts.

3.7 Evaluation metrics and performance indicators

The formative adaptive assessment system was evaluated using four complementary performance dimensions: estimation accuracy, measurement reliability, convergence stability, and item bank behavior [25]. Together, these dimensions provide a multidimensional validation framework that captures not only psychometric performance but also the interpretability, stability, and practical usability of the system in formative learning contexts. Estimation accuracy was assessed using conditional bias (CBIAS), conditional root mean squared error (CRMSE), and the Pearson correlation coefficient between true ability values θ_i and their corresponding estimates $\hat{\theta}_i$ [10]. Measurement reliability was evaluated using information-based precision derived from IRT, with the standard error of measurement. Convergence stability was examined by tracking the evolution of ability estimates $\hat{\theta}_{i,t}$ and their associated standard errors $SE_{i,t}$ across successive item administrations. This analysis provides insight into the speed and consistency with which the adaptive process yields stable proficiency estimates suitable for formative interpretation. In parallel, item bank behavior was analyzed using item exposure rates and test overlap indicators to evaluate the distribution, equity, and long-term sustainability of item usage during adaptive administration [24]. Collectively, this evaluation strategy reflects a pedagogy-first approach to adaptive assessment validation. Rather than treating performance metrics as purely technical indicators, the analyses were systematically interpreted in relation to formative assessment purposes, instructional decision-making, and curriculum alignment. This perspective ensures coherence between the system's pedagogical objectives and its evaluation framework and supports interpretations that are meaningful and actionable for engineering educators and educational designers.

4 RESULTS

This section reports the results of the simulation-based evaluation of the proposed formative adaptive assessment framework for engineering education. These results provide an integrated view of both the measurement quality and the pedagogical relevance of the proposed formative adaptive assessment framework in engineering education.

4.1 Overall performance of the formative adaptive assessment system

As shown in Table 2, the overall performance of the adaptive assessment system was evaluated using standard indicators of estimation quality. Results indicate a low conditional bias (CBIAS = 0.1127), suggesting a limited tendency toward systematic over- or underestimation of learners' proficiency levels. The CRMSE = 0.4738 reflects a moderate dispersion of estimation errors, consistent with the complexity of adaptive assessment based on a three-parameter IRT model. A strong correlation between true and estimated proficiency values ($r = 0.9119$) confirms the system's capacity to recover meaningful variability in learner performance. In addition, the mean standard error of measurement (SEM = 0.4065) and internal reliability ($\alpha = 0.8456$) exceed commonly accepted thresholds for individual-level interpretation. From a pedagogical perspective, these results support the use of the generated proficiency estimates for formative feedback, learning monitoring, and instructional decision-making in engineering education contexts [26].

Table 2. Performance of the formative adaptive assessment system

Metric	Performance Indicator	Value
Accuracy	CBIAS	0.1127
	CRMSE	0.4738
	Correlation	0.9119
	Mean Standard Error	0.4065
	Reliability	0.8456

The relationship between true and estimated proficiency levels is illustrated in Figure 5. Estimated values closely align with the ideal diagonal ($y = x$), with a slight regression-to-the-mean effect, a well-documented phenomenon in adaptive testing [27]. Increased dispersion at the lower and upper ends of the proficiency continuum indicates reduced estimation precision for extreme profiles, an issue of particular relevance for equitable formative assessment.

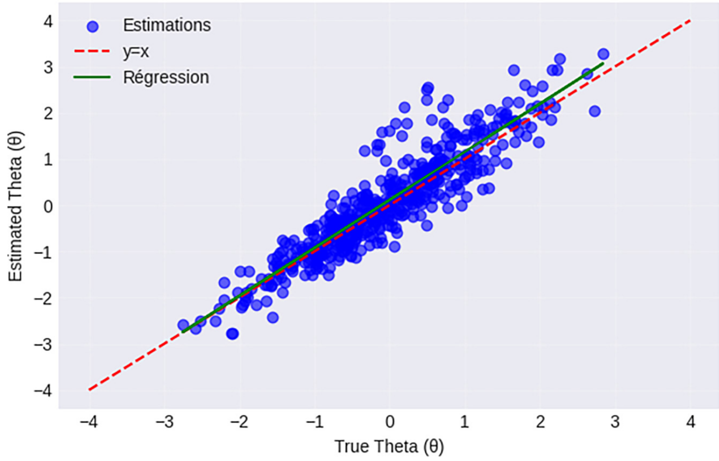

Fig. 5. Relationship between true and estimated proficiency levels in the formative adaptive assessment system

4.2 Item bank utilization and fairness considerations

Analysis of item exposure reveals uneven utilization of the item bank during adaptive administration. The overall test overlap rate (T = 26.5%) indicates that a relatively small subset of highly informative items was frequently selected across simulated administrations. Several items reached exposure rates close to 100%, raising potential concerns related to test security, fairness, and long-term sustainability.

Table 3. Item exposure and test overlap indicators

Metric	Performance Indicator	Value
Exposure	Test Overlap Rate (T)	26.5%
	Exposure Variance	0.033420

As shown in Table 3, exposure variance remains relatively low (0.0334), confirming that adaptive selection concentrated on a limited set of items. It should be noted that the 20% exposure threshold was defined as a target rather than an enforced constraint. The absence of explicit exposure control mechanisms explains why this objective was not achieved under the current configuration [34]. From an engineering education perspective, these findings highlight the importance of item bank enrichment and exposure control strategies to ensure fair access to assessment opportunities and to preserve the pedagogical and operational viability of formative adaptive assessment systems deployed at scale.

4.3 Estimation accuracy and learning regulation across proficiency levels

Estimation accuracy varies systematically across proficiency levels. Mean absolute error increases from low-ability profiles (MAE = 0.248) to medium (MAE = 0.352) and high-ability profiles (MAE = 0.400). Error distributions are more tightly centered around zero for lower-ability learners, indicating higher estimation precision, while greater dispersion is observed for medium- and high-ability profiles Figure 6.

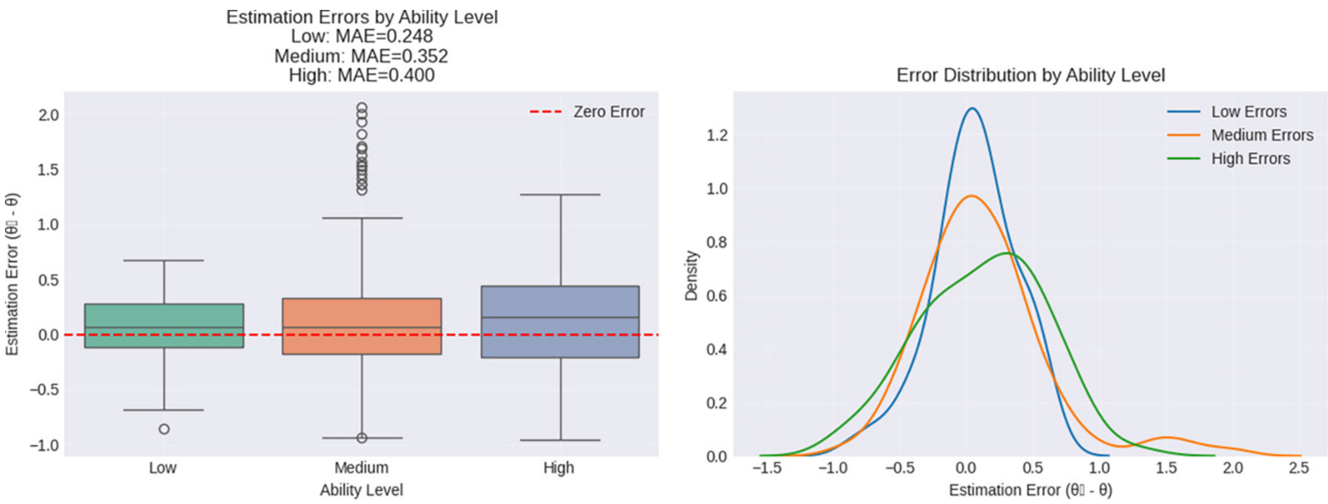

Fig. 6. Distribution of estimation errors across proficiency levels in the adaptive assessment system

The convergence of the standard error of measurement also differs across proficiency levels (Figure 7). Low-ability profiles exhibit faster convergence, with rapid reductions in standard error during the early stages of the adaptive test. In contrast, medium- and high-ability profiles show slower and incomplete convergence, and none of the groups reach the target precision threshold (SE = 0.25). From a formative assessment perspective, this pattern indicates that the system supports reliable learning regulation and diagnostic feedback for the majority of learners, but that additional item development targeting higher proficiency levels is required to ensure equitable formative feedback across the full competency continuum.

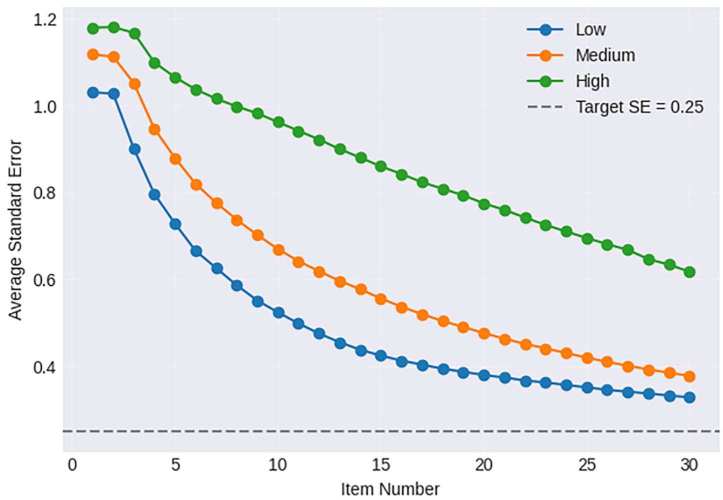

Fig. 7. Convergence of measurement precision across successive item administrations

4.4 Effects of test length on measurement precision and formative use

Analysis of test length effects confirms that measurement information is predominantly concentrated around average proficiency levels (θ ≈ 0). As the number of administered items increases, test information rises and the standard error of measurement decreases across most of the ability continuum (Figure 8). However, these gains in precision are unevenly distributed. For learners with very low or very high proficiency levels, the standard error frequently remains above the target threshold

(SE = 0.30), even with longer tests. This indicates diminishing returns of increased test length for extreme profiles and reinforces the conclusion that improvements in measurement precision are primarily achieved where item information is densest.

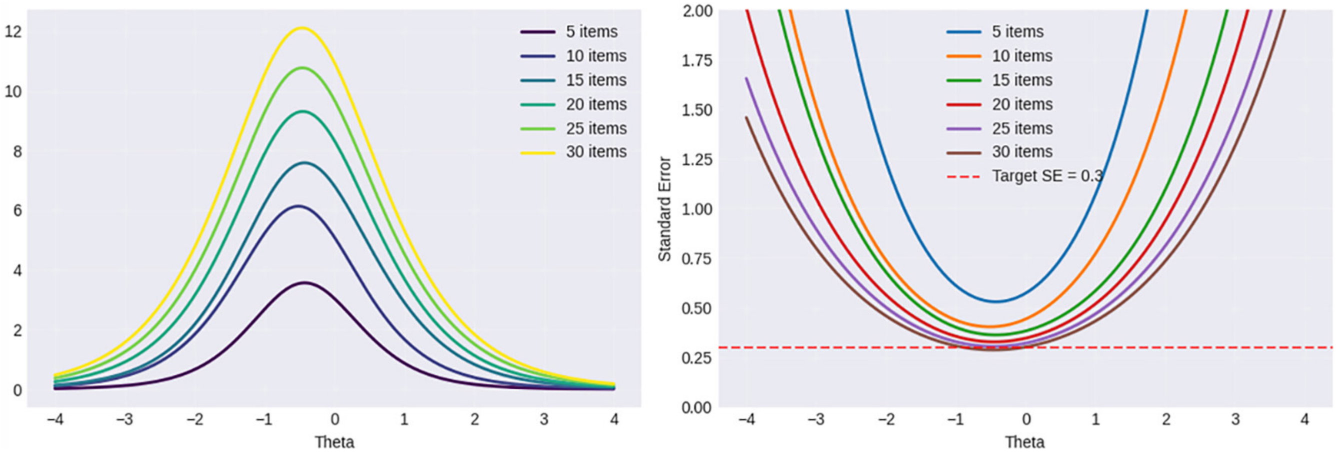

Fig. 8. Effect of test length on measurement precision across proficiency levels

Although the Bayesian diagnostic component prioritizes under-assessed competencies, it cannot fully compensate for structural imbalances in item bank coverage. These results underscore the pedagogical importance of curriculum-driven item development to support formative assessment across diverse learner profiles.

4.5 Ability estimation trajectories and implications for formative feedback

Figure 9 illustrates the evolution of ability estimates across successive item administrations for three representative learner profiles. For the low-ability profile ($\theta = -2$), initial underestimation is followed by gradual convergence toward the true value after approximately 15–20 items. The medium-ability profile ($\theta = 0$) shows moderate early fluctuations and progressively stabilizes. In contrast, the high-ability profile ($\theta = 2$) exhibits larger oscillations and slower convergence, with estimates remaining above the true value for a substantial portion of the test.

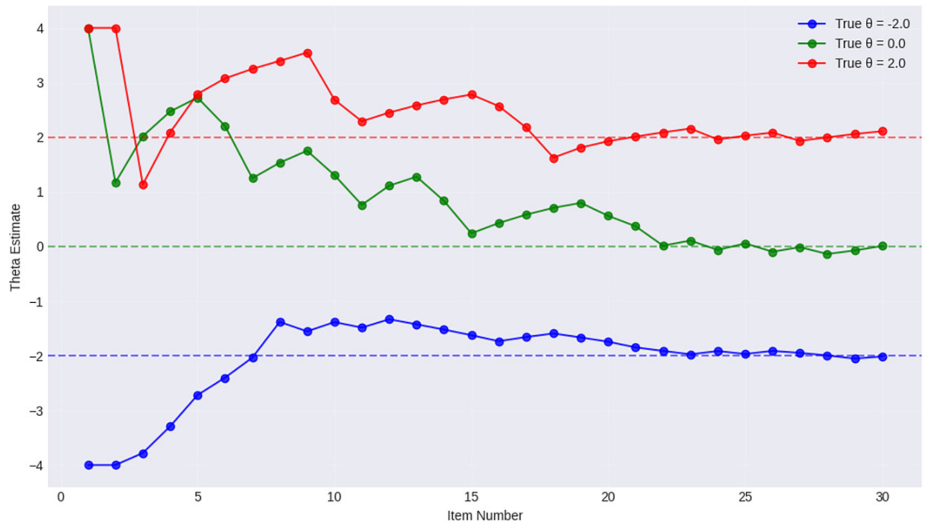

Fig. 9. Convergence of ability estimates during adaptive assessment

These trajectories illustrate the sensitivity of early estimates to initial conditions and item availability, as well as the slower stabilization observed for higher-ability learners. From an engineering education perspective, this pattern highlights the need for balanced item bank design to ensure that formative adaptive assessment supports reliable feedback, learning regulation, and instructional decision-making across all proficiency levels.

5 DISCUSSION

This study contributes to engineering pedagogy research by reframing adaptive assessment as a formative pedagogical framework rather than as a testing technology. The integration of computerized adaptive testing, competency-based diagnosis, feedback mechanisms, and human instructional decision-making positions the proposed system within the broader domain of pedagogical engineering. Rather than advancing educational measurement per se, the study demonstrates how adaptive assessment can be embedded within instructional ecosystems to support learning regulation, personalized pathways, and evidence-informed teaching. This integrative perspective responds directly to ongoing calls in engineering education research for pedagogically grounded uses of AI that enhance, rather than constrain, instructional practice [28].

5.1 Pedagogical implications for adaptive formative assessment in engineering education

The results of this study indicate that adaptive assessment, when deliberately designed as a formative pedagogical process, can meaningfully reshape assessment practices in engineering education. Rather than functioning primarily as a measurement technology, the proposed system operates as a learning-oriented assessment framework that supports teaching, learning, and instructional decision-making. In doing so, it addresses a long-standing challenge in engineering education: moving from assessment of learning toward assessment for and as learning within complex curricula and heterogeneous student populations. The strong association between true and estimated proficiency levels ($r = 0.912$) provides a necessary foundation for pedagogical confidence in the system's outputs. For instructors, such reliability is a prerequisite for integrating adaptive assessment results into everyday teaching practice. Crucially, however, the pedagogical value of the system lies not in the accuracy of a single score but in the way assessment information is structured and interpreted. By embedding proficiency estimates within a competency-oriented diagnostic framework, the system supports formative practices that make learning difficulties, prerequisite gaps, and recurring misconceptions visible elements that are often obscured by traditional summative assessments or global scores [10].

From the perspective of learning regulation, the findings underscore the role of formative adaptive assessment in supporting both learner self-regulation and instructor-led pedagogical regulation. The adaptive feedback trajectories generated by the system provide learners with timely signals about their progress, encouraging reflection, strategy adjustment, and more deliberate engagement with learning tasks. Simultaneously, instructors gain access to interpretable indicators that support responsive instructional adjustment. In engineering education, where learning frequently depends on cumulative conceptual understanding, such feedback mechanisms are particularly important for preventing persistent misunderstandings and supporting sustained learning progression. These findings align with theoretical perspectives that conceptualize formative assessment as an active driver of learning rather than a retrospective measurement activity [14]. Another key pedagogical contribution concerns the balance between personalization and assessment burden. By adaptively selecting only the most informative assessment interactions, the system reduces unnecessary testing while preserving sufficient information to support formative interpretation. This is especially relevant in engineering programs, where students often experience high assessment loads. At the same time, the system enhances pedagogical usability by providing instructors with concise, competency-focused feedback rather than extensive raw performance data. In this sense, adaptivity serves pedagogical clarity as much as efficiency. Finally, the competency-based diagnostic structure directly addresses a limitation of many adaptive testing approaches: reliance on a single global proficiency indicator. By decomposing performance into interpretable competency dimensions, the system supports pedagogically meaningful decisions such as targeted remediation, differentiated pacing, and the redesign of learning activities aligned with specific competencies. This level of diagnostic granularity is particularly well aligned with engineering education, where professional competencies are inherently multidimensional and closely connected to curriculum structures and accreditation frameworks.

5.2 Implications for instructional decision-making and faculty practice

For engineering faculty, the findings highlight the potential of formative adaptive assessment systems to function as pedagogical decision-support tools rather than as automated instructional authorities. The indicators produced by the system competency profiles, proficiency trends, and item-level patterns can be directly mobilized to inform instructional planning, content sequencing, and curriculum monitoring. At the course level, patterns of learner performance can reveal mismatches between instructional sequencing and students' prerequisite knowledge, prompting targeted adjustments to lectures, tutorials, or laboratory activities. At the programmed level, aggregated diagnostic information can support curriculum review by identifying gaps, redundancies, or misalignments across courses, thereby contributing to continuous quality enhancement processes in engineering education [29]. Importantly, the system is explicitly positioned as a diagnostic companion to instructors. While adaptive algorithms structure and synthesize assessment information, pedagogical interpretation and instructional decisions remain the responsibility of faculty. This positioning reflects a central principle of contemporary engineering pedagogy: AI supports professional judgment, it does not replace it. By reducing the cognitive and administrative burden associated with large-scale assessment, the system allows instructors to devote greater attention to instructional design, student support, and pedagogical innovation, reinforcing assessment as an integral component of teaching practice rather than a peripheral administrative task.

5.3 Assessment quality, fairness, and curriculum alignment

Beyond immediate instructional implications, the results raise broader pedagogical questions related to assessment quality, fairness, and curriculum alignment in engineering education. The concentration of assessment information around average proficiency levels, together with reduced precision at the extremes, should not be interpreted solely as technical limitations. Rather, they reflect curricular and pedagogical choices embedded in item bank design and learning outcome representation. In particular, reduced precision for novice and advanced learners highlights the need for curriculum-aligned item development that adequately represents the full range of intended learning outcomes. Engineering curricula often prioritize intermediate problem-solving competencies, which can lead to underrepresentation of foundational knowledge and advanced integrative skills in assessment instruments. These findings reinforce the view that assessment quality depends as much on curriculum design as on adaptive algorithms. Similarly, item exposure patterns raise sustainability and fairness concerns that are fundamentally pedagogical. Overexposure of certain items reflects imbalances in how competencies are represented within the assessment design. Addressing these challenges requires a sustained pedagogical commitment to balanced curriculum representation, supported by systematic item development and explicit exposure control strategies [30]. From this perspective, fairness in adaptive assessment should be understood not merely as a statistical property but as an instructional and curricular responsibility.

5.4 Limitations and future directions

The limitations of the present study should be interpreted strategically and pedagogically. First, the evaluation was based on simulation-based data, which, while appropriate for initial system validation, does not fully capture the complexity of real classroom dynamics. Second, the item bank coverage was uneven across competency levels, reflecting common curricular constraints in engineering education. Finally, the study focused on a single disciplinary context, which limits immediate generalization. Future research should therefore prioritize curriculum-driven item enrichment, particularly through the use of AI-supported automatic item generation aligned with engineering competency frameworks [31]. Longitudinal classroom deployments are also needed to examine how adaptive formative assessment influences learning trajectories, instructional practices, and student self-regulation over time. Integrating adaptive assessment with learning analytics and faculty professional development initiatives represents a particularly promising direction for supporting sustainable pedagogical change [28], [29], [32]. As a key avenue for future work, future research should examine how such formative adaptive systems are appropriated by instructors and integrated into everyday engineering teaching practices, thereby ensuring that technological innovation remains firmly anchored in pedagogical purpose.

6 CONCLUSION

This paper presented a formative adaptive assessment framework for engineering education, designed to align computerized adaptive assessment with curriculum objectives, competency-based diagnosis, and instructional decision-making. By integrating a curriculum-aligned item bank, an IRT-based adaptive testing engine, and diagnostic modeling, the proposed system reframes adaptive assessment as a pedagogical tool rather than a purely measurement-oriented technology. The empirical calibration and simulation-based evaluation demonstrate that the framework can produce reliable, interpretable, and pedagogically meaningful proficiency indicators suitable for formative use. The strong correspondence between true and estimated proficiency levels confirms the system's capacity to support instructional interpretation, while competency-oriented diagnostics extend adaptive assessment beyond global scores. These features enable instructors to identify learning difficulties, monitor progress, and support learning regulation across heterogeneous learner populations commonly found in engineering programs. Beyond measurement performance, the findings emphasize that the effectiveness and fairness of adaptive assessment are closely tied to curriculum alignment and item bank design. Reduced precision at the extremes of the proficiency continuum and imbalances in item exposure highlight that adaptive algorithms cannot compensate for gaps in curriculum representation. As such, assessment quality in engineering education depends as much on pedagogically grounded item development as on adaptive control mechanisms. Importantly, the framework positions adaptive assessment as a decision-support system that enhances, rather than replaces, faculty judgment. By reducing assessment burden while improving the interpretability of assessment data, the system supports evidence-informed teaching, personalized learning pathways, and continuous curriculum improvement. This positioning aligns with core principles of engineering pedagogy and responds to current calls for pedagogically grounded applications of AI in education. Overall, this study contributes to engineering education research by demonstrating how computerized adaptive assessment can be embedded within a formative, curriculum-sensitive pedagogical framework. Future research should prioritize longitudinal classroom implementations, curriculum-driven item enrichment, and faculty appropriation of adaptive feedback tools to ensure that adaptive assessment technologies sustainably support teaching, learning, and educational quality in engineering education.

7 REFERENCES

[1] Y. Laaziz, G. Chemsi, and M. Radid, "The influence of e-assessment on students' cognitive engagement in higher education," *International Journal of Engineering Pedagogy (iJEP)*, vol. 14, no. 4, pp. 54–67, 2024. https://doi.org/10.3991/ijep.v14i4.48293

[2] K. Smahi, O. Labouidya, and K. E. Khadiri, "Enhancing online assessment quality in higher education: The design of moodle plug-in for personalised exam revision (PER)," *International Journal of Engineering Pedagogy (iJEP)*, vol. 15, no. 5, pp. 127–140, 2025. https://doi.org/10.3991/ijep.v15i5.55055

[3] M. E. Msayer, E.-S. Aoula, and B. Bouihi, "Artificial intelligence in computerized adaptive testing to assess the cognitive performance of students: A systematic review," in *2024 International Conference on Intelligent Systems and Computer Vision (ISCV)*, 2024, pp. 1–8. https://doi.org/10.1109/ISCV60512.2024.10620092

[4] K. Sodoké, G. Raîche, and R. Nkambou, "La plateforme d'évaluation adaptative des apprentissages: PersonFit," in *Environnements Informatiques pour l'Apprentissage Humain*, 2007. https://hal.science/hal-00161592v1

[5] Q. Liu *et al.*, "Survey of computerized adaptive testing: A machine learning perspective," *arXiv preprint arXiv:2404.00712*, 2024. https://doi.org/10.48550/arXiv.2404.00712

[6] P. T. D. Thuy, N. T. Van, N. H. Trang, N. M. Thuong, V. T. T. Hien, and V. Q. Chung, "Designing an AI-supported formative assessment model for pre-service mathematics teacher self-study in Vietnam," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 19, no. 22, pp. 50–68, 2025. https://doi.org/10.3991/ijim.v19i22.57723

[7] D. Spenassato, A. C. Bornia, and R. Tezza, "Computerized adaptive testing: A review of research and technical characteristics," *IEEE Latin America Transactions*, vol. 13, no. 12, pp. 3890–3898, 2015. https://doi.org/10.1109/TLA.2015.7404924

[8] E. Tsutsumi, R. Kinoshita, and M. Ueno, "Deep item response theory as a novel test theory based on deep learning," *Electronics*, vol. 10, no. 9, p. 1020, 2021. https://doi.org/10.3390/electronics10091020

[9] M. Kaplan, J. de la Torre, and J. R. Barrada, "New item selection methods for cognitive diagnosis computerized adaptive testing," *Appl. Psychol. Meas.*, vol. 39, no. 3, pp. 167–188, 2015. https://doi.org/10.1177/0146621614554650

[10] Y. Choi and C. McClenen, "Development of adaptive formative assessment system using computerized adaptive testing and dynamic Bayesian networks," *Applied Sciences*, vol. 10, no. 22, p. 8196, 2020. https://doi.org/10.3390/app10228196

[11] L. Burigana, "Bayesian networks and knowledge structures in cognitive assessment: Remarks on basic comparable aspects," *Journal of Mathematical Psychology*, vol. 123, p. 102875, 2024. https://doi.org/10.1016/j.jmp.2024.102875

[12] N. Thompson and D. Weiss, "A framework for the development of computerized adaptive tests," *Practical Assessment, Research, and Evaluation*, vol. 16, no. 1, 2019. https://doi.org/10.7275/wqzt-9427

[13] M. von Davier, "Automated item generation with recurrent neural networks," *Psychometrika*, vol. 83, no. 4, pp. 847–857, 2018. https://doi.org/10.1007/s11336-018-9608-y

[14] J. B. Bjorner, C.-H. Chang, D. Thissen, and B. B. Reeve, "Developing tailored instruments: Item banking and computerized adaptive assessment," *Qual. Life Res.*, vol. 16, pp. 95–108, 2007. https://doi.org/10.1007/s11136-007-9168-6

[15] J. Sharpnack, P. Mulcaire, K. Bicknell, G. LaFlair, and K. Yancey, "AutoIRT: Calibrating item response theory models with automated machine learning," *arXiv preprint arXiv:2409.08823*, 2024. https://doi.org/10.48550/arXiv.2409.08823

[16] S. E. Stemler and A. Naples, "Rasch measurement v. Item response theory: Knowing when to cross the line," *Practical Assessment, Research, and Evaluation*, vol. 26, no. 1, 2021. https://doi.org/10.7275/v2gd-4441

[17] R. Ostini and M. Nering, *Polytomous Item Response Theory Models*. Thousand Oaks, CA: SAGE Publications, Inc., 2006. https://doi.org/10.4135/9781412985413

[18] Y. Zhuang, Q. Liu, Z. Huang, Z. Li, S. Shen, and H. Ma, "Fully adaptive framework: Neural computerized adaptive testing for online education," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, pp. 4734–4742, 2022. https://doi.org/10.1609/aaai.v36i4.20399

[19] E. V. Chumakova, T. A. Chernova, Y. A. Belyaeva, D. G. Korneev, and M. S. Gasparian, "Use of neural networks in the adaptive testing system," *IJACSA*, vol. 13, no. 5, 2022. https://doi.org/10.14569/IJACSA.2022.0130504

[20] T. Song *et al.*, "Novel application of deep learning for adaptive testing based on long short-term memory," in *2019 IEEE 37th VLSI Test Symposium (VTS)*, 2019, pp. 1–6. https://doi.org/10.1109/VTS.2019.8758628

[21] M. Phankokkruad and K. Woraratpanya, "An automated decision system for computer adaptive testing using genetic algorithms," in *2008 Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, 2008, pp. 655–660. https://doi.org/10.1109/SNPD.2008.118

[22] M. Ueno and P. Songmuang, "Computerized adaptive testing based on decision tree," in *2010 10th IEEE International Conference on Advanced Learning Technologies*, 2010, pp. 191–193. https://doi.org/10.1109/ICALT.2010.58

[23] M. Imran, N. Almusharraf, M. S. Abdellatif, and M. Y. Abbasova, "Artificial intelligence in higher education: Enhancing learning systems and transforming educational paradigms," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 18, no. 18, pp. 34–48, 2024. https://doi.org/10.3991/ijim.v18i18.49143

[24] V. M. G. Jatobá, J. S. Farias, V. Freire, A. S. Ruela, and K. V. Delgado, "ALICAT: A customized approach to item selection process in computerized adaptive testing," *Journal of the Brazilian Computer Society*, vol. 26, no. 1, p. 4, 2020. https://doi.org/10.1186/s13173-020-00098-z

[25] J. R. Barrada, J. Olea, V. Ponsoda, and F. J. Abad, "Incorporating randomness in the Fisher information for improving item-exposure control in CATs," *Br. J Math. Stat. Psychol.*, vol. 61, pp. 493–513, 2008. https://doi.org/10.1348/000711007X230937

[26] J. Wang, Y. Xie, Z. Feng, and J. Li, "Psychometric properties of computerized adaptive testing for chronic obstructive pulmonary disease patient-reported outcome measurement," *Health Qual. Life Outcomes*, vol. 22, p. 73, 2024. https://doi.org/10.1186/s12955-024-02291-6

[27] R. K. Hambleton and H. Swaminathan, *Item Response Theory*. Netherlands: Springer Netherlands, 1985. https://doi.org/10.1007/978-94-017-1988-9

[28] D. Kazimova, G. Tazhigulova, G. Shraimanova, A. Zatyneyko, and A. Sharzadin, "Transforming university education with AI: A systematic review of technologies, applications, and implications," *International Journal of Engineering Pedagogy (iJEP)*, vol. 15, no. 1, pp. 4–24, 2025. https://doi.org/10.3991/ijep.v15i1.50773

[29] A. Klein, J. Müller, N. Link, and R. Schaffer, "Assessment and level modelling in fundamentals of electrical engineering," *International Journal of Engineering Pedagogy (iJEP)*, vol. 14, no. 6, pp. 37–53, 2024. https://doi.org/10.3991/ijep.v14i6.47865

[30] M. L. Stocking and C. Lewis, "Methods of controlling the exposure of items in CAT," in *Computerized Adaptive Testing: Theory and Practice*, W. J. van der Linden and G. A. W. Glas, Éds., Dordrecht: Springer Netherlands, 2000, pp. 163–182. https://doi.org/10.1007/0-306-47531-6_9

[31] Y. Song, J. Du, and Q. Zheng, "Automatic item generation for educational assessments: A systematic literature review," *Interactive Learning Environments*, vol. 33, no. 9, pp. 5386–5405, 2025. https://doi.org/10.1080/10494820.2025.2482588

[32] M. Bellaj, A. B. Dahmane, S. Boudra, and M. L. Sefian, "Big data analytics in higher education: A new adaptive learning analytics model integrating traditional approaches," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 18, no. 6, pp. 24–39, 2024. https://doi.org/10.3991/ijim.v18i06.46289

[33] I. Azaiz, O. Deckarm, and S. Strickroth, "AI-Enhanced auto-correction of programming exercises: How effective is GPT-3.5?" *International Journal of Engineering Pedagogy (iJEP)*, vol. 13, no. 8, pp. 67–83, 2023. https://doi.org/10.3991/ijep.v13i8.45621

[34] S. Lim and S. W. Choi, "Item exposure and utilization control methods for optimal test assembly," *Behaviormetrika*, vol. 51, no. 1, pp. 125–156, 2024. https://doi.org/10.1007/s41237-023-00214-1

8 AUTHORS

Mohamed El Msayer received a Master's degree in Data Engineering and Software Development. He is currently a Ph.D. student in computer science and artificial intelligence with the Laboratory 2IACS, ENSET Mohammedia, Hassan II University of Casablanca. His doctoral research lies at the intersection of psychometrics, computerized adaptive testing, and artificial intelligence, with a particular emphasis on the adaptive assessment of cognitive competencies. His work focuses on theoretical and methodological advances in item response Theory, including unidimensional and multidimensional modeling, item calibration, ability estimation, and adaptive item selection strategies (E-mail: mohamed.elmsayer-etu@etu.univh2c.ma).

Bouchra Bouihi has completed engineering in Computer Science from the National School of Applied Sciences in 2014 and the Ph.D. degree in Computer Science from the Faculty of Science and Technology, Hassan 1st University, Settat, Morocco, in 2019. She is currently an Affiliate Professor with the Department of Mathematics and Computer Science and a Research Member of the 2IACS Laboratory, ENSET Mohammedia, University Hassan II, Casablanca. Her research interests include artificial intelligence, with specific interests in machine learning, deep learning, and ontology engineering. Her work delves into solving real-world problems, including the education field, by using AI models.

Abdelmajid Bousselham received the Ph.D. degree in Computer Science and Artificial Intelligence. He is currently a Professor and a researcher at ENSET Mohammedia and a member of the Laboratory of Computer Science, Artificial Intelligence, and Cybersecurity (2IACS), Hassan II University of Casablanca. His research interests encompass artificial intelligence and deep learning, with particular emphasis on biomedical image analysis, medical image segmentation, multimodal learning, and computer-aided diagnosis systems. His work is devoted to the development of advanced AI-based models aimed at addressing real-world healthcare challenges, including disease detection and medical decision support.

Essaadia Aoula has completed Doctor of Education. Her research focuses on the integration of artificial intelligence in education, with particular emphasis on personalized learning, student performance enhancement, academic dropout prevention, and automated assessment systems. Her work aims to improve the quality of digital and blended learning environments and to strengthen employability across both initial and continuing education.

Adel Deraoui received the Ph.D. degree in Computer Science, with a specialization in educational technology and pedagogy. He is currently affiliated with the Casablanca–Settat Regional Center for Education and Training, Morocco. His interests include educational technologies, digital pedagogy, and the integration of information and communication technologies in education.