International Research Journal of Multidisciplinary Scope (IRJMS), 2026;7(3):923-934

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2026.v07i03.010127

Adaptive Hybrid Model for Knowledge Assessment Integrating IRT Estimation, Temporal Dynamics and Semantic Analysis

Medetova Kunduz Muratovna[1*], Nadira Xalimovna Latipova[2], Larisa Timofeevna Marisheva[2], Shavkat Mugaveevich Ravilov[3]

[1]Department for Cooperation with International Ranking and Accreditation Agencies, Tashkent National Pedagogical University named after Nizami, Tashkent, Uzbekistan, [2]Department of Systems and Applied Programming, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent, Uzbekistan, [3]Department of Mathematics and Informatics, Tashkent Branch of Gubkin Russian State University of Oil and Gas (National Research University), Tashkent, Uzbekistan.

*Corresponding Author's Email: medetovakm@gmail.com

Abstract

This paper proposes an adaptive hybrid model for knowledge assessment integrating latent ability estimation based on Item Response Theory (IRT), temporal dynamics of learner behaviour and semantic analysis of open-ended responses within a unified computational framework. Conventional assessment systems predominantly rely on static test scores and closed-form items, which limits their ability to capture behavioural patterns and qualitative aspects of learner performance. The proposed model refines the latent knowledge parameter during testing by incorporating response-time distributions modelled through a log-normal framework and stable behavioural indicators derived from user interaction data. Open-ended answers are processed through semantic and structural feature extraction using pre-trained language model embeddings, producing a normalized auxiliary score that contributes to the overall competence estimation. The methodological framework combines probabilistic latent trait modelling, temporal analytics and semantic representation techniques to ensure dynamic and multidimensional evaluation. A weighted hybrid integration strategy aggregates IRT-based ability estimates, temporal adjustment indices and semantic similarity scores into a single latent competence value updated iteratively through an adaptive learning rule. The hybrid structure enables the system to adapt to performance fluctuations and heterogeneous response formats within a unified computational scheme. Unlike existing adaptive testing and AI-driven scoring systems that address a single data modality, the proposed framework simultaneously integrates three complementary information sources. Experimental validation conducted in a corporate environment demonstrated an 11.7% improvement in assessment accuracy and a 4.3-minute reduction in testing time compared to conventional non-adaptive approaches. The proposed approach provides a scalable and computationally grounded solution for adaptive testing environments, digital education platforms and personnel assessment systems requiring accurate, personalized and data-driven evaluation mechanisms.

Keywords: Adaptive Assessment, Automated Knowledge Assessment, Item Response Theory, Latent Ability Estimation, Semantic Analysis, Temporal Dynamics.

Introduction

Digital learning technologies have undergone rapid development over the past decade, leading to a growing demand for assessment systems that can adapt to a wide range of learners and provide results that are both accurate and meaningful. As users interact with online testing platforms, they leave behind a large amount of behavioural and textual data, which traditional accuracy-based scoring models do not fully utilize. Because of this, researchers increasingly point out the need for diagnostic approaches that move beyond simple correctness scoring and incorporate multiple dimensions of user behaviour. Classical psychometric models, particularly Item Response Theory (IRT), have had a major influence on the field and remain widely used due to their ability to estimate latent knowledge levels with statistical precision (1, 2). While these models form the backbone of many large-scale assessments, several studies have shown that accuracy alone may not capture important cognitive and behavioural differences between test-takers (3). In digital learning environments, students interact with tasks in ways that reveal much more than just right or wrong answers. One important source of information is response time, which reflects how

This is an Open Access article distributed under the terms of the Creative Commons Attribution CC BY license (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution and reproduction in any medium, provided the original work is properly cited.

(Received 15th January 2026; Accepted 09th June 2026; Published 16th July 2026)

quickly and confidently users make decisions. Early research demonstrated that analyzing both accuracy and response time gives a more complete understanding of test performance and cognitive processes (4). Later studies confirmed that timing patterns often correspond to levels of effort, familiarity with the material and the strategies used to approach different items (5). For example, learners might rush through easy items but slow down on complex ones, or show stable time patterns when they understand the material well. These behavioral signals have been shown to improve the stability of latent trait estimates, especially in adaptive testing settings where decisions about the next item depend on previous performance (6, 7). Because digital systems automatically record precise timings, response-time modelling has become one of the key directions in modern assessment research.

In addition to behavioral indicators, open-ended responses provide another rich source of information about a learner's knowledge and reasoning. Traditionally, these responses were scored manually, but advances in natural language processing have allowed automated systems to identify meaning, structure and depth of understanding within free-text answers. Early approaches focused on lexical matching and surface features, but more advanced models consider semantic similarity, sentence structure and key conceptual elements (8). Studies have shown that semantic and structural features of student writing often correlate strongly with deeper comprehension and higher-level thinking skills (9). As a result, several automated scoring systems have been developed for large-scale educational assessments, demonstrating that semantic analysis can complement traditional scoring approaches and reveal aspects of learning that are not visible in closed-ended items (10, 11). However, despite these advances, semantic scoring and psychometric modelling are still commonly treated as separate methodologies, each applied independently depending on the type of assessment task.

This separation leads to a situation where existing assessment systems frequently rely on a single data modality - either correctness, timing, or semantic features - rather than combining them. A growing number of researchers argue that this limits the potential accuracy and diagnostic value of digital testing. Several recent attempts have been made to integrate behavioral and textual data, but these studies generally examine narrow cases or theoretical models that have not yet been fully implemented in adaptive testing environments (12). This leaves an important methodological gap: although multiple sources of information are available, there is no widely adopted computational framework that unifies latent modelling, temporal behaviour and semantic processing into a single coherent assessment mechanism. Existing adaptive testing systems such as computerized adaptive testing platforms typically focus on IRT-based item selection without incorporating semantic or temporal features. AI-driven automated scoring models address textual response quality independently, without linkage to a latent trait estimation pipeline. Multimodal assessment frameworks proposed in the learning analytics literature rarely operate in real time and do not integrate adaptive item selection. The proposed model addresses all three limitations simultaneously by embedding probabilistic estimation, temporal modelling and semantic analysis within a unified adaptive testing architecture.

Another challenge is the diversity of user behaviour in digital settings. People differ in their pacing strategies, in how they allocate time across tasks, in how they interpret written instructions and in how they express ideas in text. Cognitive psychology has repeatedly shown that such differences influence performance and can lead to inconsistent or biased estimates when models rely on only one type of data (13). For example, fast test-takers may be unfairly judged as careless when only accuracy is considered, while slow and careful respondents might receive unnecessarily low estimates if time is not taken into account. Similarly, two open-ended responses of equal length may differ significantly in conceptual quality, which can only be captured through semantic analysis. This reinforces the need for hybrid assessment models capable of integrating multiple forms of evidence. The increasing overlap between computational modelling, psychometrics and machine learning has opened possibilities for new approaches that can process heterogeneous data in a unified way. Several works highlight that advanced computational tools enable the integration of behavioural, linguistic and statistical signals, providing a richer basis for decision-making in adaptive systems (14). Yet most existing approaches still focus on one primary modality, often because integrating different types of data requires more complex architectures and careful calibration. As digital learning environments continue to evolve, the need for more flexible and comprehensive assessment models becomes increasingly evident.

Given these developments, an opportunity emerges to design a hybrid assessment framework that brings together three complementary sources of information: latent trait estimation from IRT, behavioral timing patterns from response-time data and semantic representation of open-ended responses. Each of these elements contributes unique insights into user performance and their combination can support a more balanced and informative assessment. The hybrid approach offers a way to capture both the outcome of user responses and the underlying processes that lead to them, thus moving assessment closer to real-world learning behavior.

The present study builds on this idea by proposing an adaptive model that integrates the strengths of psychometric theory, behavioral modelling and semantic analysis. The model is designed for environments where user interaction data are readily available and can be processed in real time. By unifying these different sources of information, the model creates a more stable and reliable estimate of knowledge level, which can improve item selection, reduce measurement error and provide more meaningful feedback to learners. This work contributes to the field of multimodal assessment by presenting a practical and computationally grounded framework that aligns with current trends in digital education and offers new opportunities for research and application.

Methodology

Latent Ability Modeling

The first component of the proposed hybrid framework is a latent-trait estimation module grounded in the three-parameter logistic Item Response Theory model. This formulation has been shown to provide a stable link between item characteristics and examinee proficiency, especially in adaptive environments where item selection depends on real-time ability updates.

The probability of a correct response is expressed as Equation [1]:

$$P_i(\theta) = c_i + \frac{1-c_i}{1+\exp[-a_i(\theta-b_i)]} \quad [1]$$

The log-likelihood for a response vector X is expressed in Equation [2]:

$$l(\theta|X) = \sum_{i=1}^{n}\left[x_i \ln P_i(\theta) + (1-x_i)\ln\left(1-P_i(\theta)\right)\right] \quad [2]$$

Estimation of the latent ability parameter θ is performed using a Bayesian updating scheme. A normal prior distribution $\theta \sim \mathcal{N}(0,1)$ is assumed at the initial stage. After each response, the posterior distribution is updated according to Bayes' rule and the expected value of the posterior is used as the current ability estimate. This sequential updating mechanism enables stable real-time adaptation of the model during the testing process.

To enhance the robustness of the estimation, response time information is incorporated through a log-normal temporal model. Let t_i denote the response time for item i.

The distribution of t_i is modeled as shown in Equation [3]:

$$\ln t_i \sim \mathcal{N}(\mu_i, \sigma_i^2) \quad [3]$$

Where, μ_i and σ_i^2 represent the expected response speed and variability, respectively.

Deviations from typical response time patterns are used as additional behavioral indicators for refining the latent ability estimate.

The final latent ability value is obtained through the integration of probabilistic response accuracy and temporal dynamics, providing a more informative and behavior-aware estimation of learner competence compared to classical IRT formulations.

Temporal Dynamics Modeling

In adaptive testing environments, response accuracy alone is insufficient for reliable estimation of latent ability, since it does not reflect hesitation, cognitive load, or guessing behavior. Therefore, temporal characteristics of interaction with test items are incorporated as an additional behavioral modality.

For each item i, the response time t_i is modeled using a log-normal distribution with probability density function as given in Equation [4].

$$f(t_i) = \frac{1}{t_i \sigma_i \sqrt{2\pi}} \exp\left(-\frac{(\ln t_i - \mu_i)^2}{2\sigma_i^2}\right) \qquad [4]$$

Based on this representation, a normalized temporal indicator is computed as shown in Equation [5].

$$T_i = \frac{\ln t_i - \mu_i}{\sigma_i} \qquad [5]$$

The obtained value characterizes deviations from typical response behavior. The temporal adjustment of latent ability is defined in Equation [6].

$$\theta_t = \theta - \lambda T_i \qquad [6]$$

This is a regularization coefficient controlling the influence of time on the final estimate.
This approach allows the model to account for abnormal response speeds while preserving the stability of probabilistic ability estimation.

Semantic Analysis of Open-ended Responses

To assess the quality of open-ended responses, a semantic analysis module based on vector representations of text is employed. Each learner response u and the corresponding reference answer r are transformed into embedding vectors v_u and v_r, respectively, using pre-trained language models, i.e., $v_u = \mathrm{Embed}(u), v_r = \mathrm{Embed}(r)$.
The semantic similarity between the two texts is computed using cosine similarity as expressed in Equation [7]:

$$S = \frac{v_u \cdot v_r}{|v_u||v_r|} \qquad [7]$$

To ensure compatibility with the latent ability scale, the similarity score is normalized as shown in Equation [8]:

$$\theta_s = \frac{S - \min(S)}{\max(S) - \min(S)} \qquad [8]$$

This normalized semantic component reflects the conceptual proximity of the learner's answer to the reference solution and serves as a qualitative indicator of knowledge depth.

Hybrid Integration Strategy

The proposed framework integrates probabilistic accuracy, temporal behavior and semantic understanding into a unified latent competence estimate. Let θ_a denote the latent ability obtained from the IRT model, θ_t the temporal adjustment and θ_s the semantic component. The aggregated latent competence estimate is computed as a weighted combination $\theta^{=} w_1\theta_a + w_2\theta_t + w_3\theta_s$, subject to the constraint $w_1 + w_2 + w_3 = 1, w_i \geq 0$.

The final estimate is updated iteratively according to the learning rule given in Equation [9]:

$$\theta_{k+1} = \theta_k + \eta(\theta^{-}\theta_k) \qquad [9]$$

Where, η is the adaptation rate.

Such a hybrid integration mechanism ensures stable convergence of the latent ability estimate while simultaneously incorporating behavioral and semantic information into the adaptive testing process.

Adaptive Item Selection Based on Information Criterion

Adaptive item selection is a key component of the proposed knowledge assessment system, as it directly determines the efficiency and precision of the testing process.

At each step of the adaptive cycle, the system selects the next test item based on the current estimate of the latent ability θ_k. The objective is to maximize the expected information gain while minimizing the total number of administered items, which is consistent with classical principles of computerized adaptive testing (15) and their application in adaptive learning environments (16, 17).

For each candidate item i, the Fisher information function under the three-parameter logistic IRT model is computed as shown in Equation [10].

$$I_i(\theta) = \frac{a_i^2(1-P_i(\theta))(P_i(\theta)-c_i)^2}{(1-c_i)^2 P_i(\theta)} \qquad [10]$$

Where, a_i is the discrimination parameter, c_i is the guessing parameter and $P_i(\theta)$ is the probability of a correct response expressed in Equation [1].

The next test item is selected according to the maximum information criterion given in Equation [11].

$$i^* = \arg\max_{i \in \mathcal{B}} I_i(\theta_k) \qquad [11]$$

Where, $\mathcal{B}$ denotes the current pool of available items and θ_k is the current estimate of the latent ability after k iterations.

To prevent over-exploitation of items with extremely high discrimination and to ensure content validity, an additional regularization term is introduced in the selection rule. The modified selection criterion is defined in Equation [12] as,

$$i^* = \arg\max_{i \in \mathcal{B}} (I_i(\theta_k) - \beta R_i) \qquad [12]$$

Where, R_i is a content-balancing penalty and β controls the trade-off between information gain and test diversity.

After the response to the selected item is obtained, the latent ability estimate is iteratively updated using the hybrid integration mechanism described in Section 2.4, which allows the system to continuously refine the current competence level based on incoming evidence (18). This closed-loop adaptive procedure continues until a predefined stopping condition is satisfied, such as achieving a target standard error of estimation or reaching a maximum test length. The proposed information-driven selection strategy enables a significant reduction in test length while preserving high measurement accuracy, which has been consistently confirmed by classical studies in computerized adaptive assessment (19).

System Architecture and Implementation

The proposed automated knowledge assessment system is implemented as a modular, service-oriented architecture designed to support real-time adaptive testing, multimodal data processing and continuous updating of latent competence estimates. The architectural design follows the general principles of scalable intelligent assessment platforms and data-driven educational systems widely discussed in the literature on computerized adaptive testing and educational data mining (20).

At the input layer, the system integrates multiple heterogeneous data sources, including a learning management system, a bank of test items, open-ended textual responses, monitored behavioral indicators and external data repositories. All incoming data are transmitted to the data preprocessing module, where primary filtration, normalization of temporal parameters and unification of data formats are performed. This stage is critical for ensuring the stability and reliability of the subsequent probabilistic and semantic analysis, which is consistent with general data-driven system design approaches in learning analytics.

Figure 1 illustrates the general architecture of the proposed system, including all main functional modules and data flows between them.

Access to the system is regulated through a dedicated security subsystem that includes authentication, authorization and role-based access control. Administrative functions are implemented via a separate administrator console, which provides tools for system parameter configuration, assignment of access rights and continuous monitoring and auditing of user actions. This ensures both data protection and full traceability of all operations performed within the platform.

Preprocessed data are stored in a centralized data repository, which serves as the core storage component of the system. From this repository, data are transferred to the analytical layer, which integrates several intelligent modules operating in parallel. This layer includes the IRT-based probabilistic assessment module, the Bayesian knowledge tracing module responsible for modeling learning dynamics and the semantic analysis module for processing open-ended responses. The parallel processing of heterogeneous information streams enables the system to capture complementary aspects of learner behavior and knowledge formation.

The outputs of the analytical modules are forwarded to the iterative integration layer, where model parameters are dynamically updated and synchronized. A dedicated registry and version control mechanism for analytical models supports continuous system learning and incremental refinement of model parameters as new data become available. The training module performs periodic recalibration of probabilistic and semantic models, while the assessment module generates updated estimates of the current knowledge level.

The adaptive item selection module receives the updated latent competence estimate and determines the next most informative test item according to the information-based criterion described in the Methods section. This mechanism ensures dynamic adaptation of the testing trajectory to individual learner characteristics and allows for a significant reduction in the total number of administered tasks without loss of measurement accuracy.

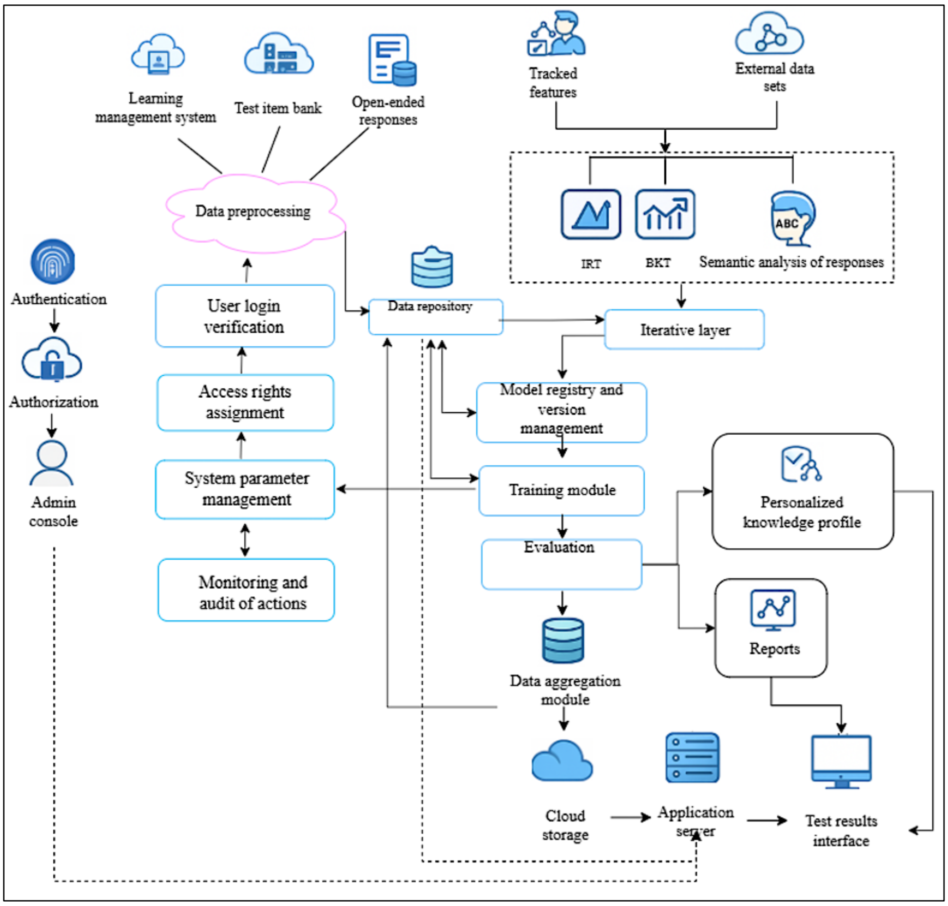

Figure 1: Architecture of the Adaptive Hybrid Knowledge Assessment System

A separate data aggregation module collects intermediate and final analytical results and transfers them to cloud storage and the application server. On the basis of the aggregated data, the system forms a personalized knowledge profile for each employee, which reflects the current level of competence, detected strengths and weaknesses and learning gaps. Using this profile, the reporting and recommendation module generates analytical reports and individualized learning recommend-dations.

The final assessment results are delivered to end users through a dedicated testing results interface that provides visual access to evaluation outcomes, dynamics of knowledge development and personalized learning recommendations. From an implementation perspective, all functional compo-nents of the system are deployed as loosely coupled services interacting through a unified data exchange interface, which ensures reliable coordination of system processes and robust data flow management (20). Such architectural organization increases system scalability, simplifies integration into existing corporate infrastructures and enables flexible functional extension without restructuring the core framework.

The proposed architecture ensures full automation of the entire knowledge assessment cycle, from initial data acquisition and preprocessing to latent competence profiling and recommendation generation. Its modular structure allows the system to evolve in accordance with organizational requirements and emerging analytical technologies, which is fully consistent with modern trends in intelligent educational and assessment platforms (21).

Experimental Setup

The proposed intelligent knowledge assessment system was experimentally evaluated in a real corporate environment within an IT company. The experimental study involved a group of employees representing different levels of professional experience and qualification. The main objective of the experiment was to assess the effectiveness of the proposed adaptive hybrid assessment framework in terms of accuracy, efficiency and user activity. The experiment was organized according to a “before–after” evaluation scheme. At the first stage, employees were assessed using a conventional non-adaptive testing approach. At the second stage, the same group of participants was evaluated using the proposed automated adaptive assessment system. This design made it possible to directly compare the performance of classical and adaptive assessment mechanisms under identical conditions.

The testing materials included both closed-form test items and open-ended questions. Response correctness, response time, error frequency and user activity indicators were recorded for each participant. Semantic analysis was applied to open-ended responses, while probabilistic and temporal characteristics were processed using the proposed hybrid modeling framework.

To ensure objectivity of the evaluation, the same test content and time constraints were used at both stages of the experiment. All collected data were aggregated and statistically processed to compute average values and relative changes in key performance indicators. Descriptive statistics were calculated for all variables, including means and standard deviations. Paired comparisons between the before- and after-implementation conditions were evaluated using a paired-samples t-test, with a significance level set at $p < 0.05$. Effect sizes were computed using Cohen’s d to quantify the practical significance of observed differences. The main evaluation metrics included assessment accuracy, average testing time, number of errors and staff activity level.

Results

The experimental evaluation of the proposed adaptive hybrid knowledge assessment system demonstrated a consistent positive effect across all key performance indicators. Quantitative results of the “before-after” comparison are summarized in Table 1.

Table 1: Comparative Results of Knowledge Assessment Before and After System Implementation

Indicator	Before Implementation	After Implementation	Change
Assessment accuracy (%)	72.4	84.1	+11.7
Average testing time (min)	28.6	24.3	−4.3
Average number of errors	2.6	1.9	−0.7
Staff activity level (%)	48	62	+14

As shown in Table 1, the implementation of the proposed system led to a substantial increase in assessment accuracy. The average accuracy improvement of 11.7% confirms that the integration of probabilistic IRT estimation, tempo-ral dynamics and semantic analysis enables a more precise and robust measurement of employee competence compared to conventional non-adaptive testing. This result indicates that the hybrid approach effectively captures both quantitative correctness and qualitative depth of understanding.

The average testing time was reduced by 4.3 minutes, which demonstrates the efficiency of the adaptive item selection mechanism. By dynamically selecting the most informative tasks based on the current latent ability estimate, the system avoids redundant items and shortens the testing trajectory without compromising measurement precision. This confirms the practical value of the information-based selection strategy implemented in the system. A noticeable decrease in the average number of errors (-0.7) was also observed after system deployment. This effect can be explained by the improved correspondence between task difficulty and individual competence level, which reduces random guessing and cognitive overload. The incorporation of response time and semantic analysis additionally contributes to the stabilization of assessment outcomes.

The staff activity level increased by 14%, indicating higher user engagement under the adaptive assessment regime. The growth in activity can be attributed to personalized testing trajectories, clearer feedback and the generation of individualized recommendations. These factors enhance user motivation and perceived usefulness of the assessment process.

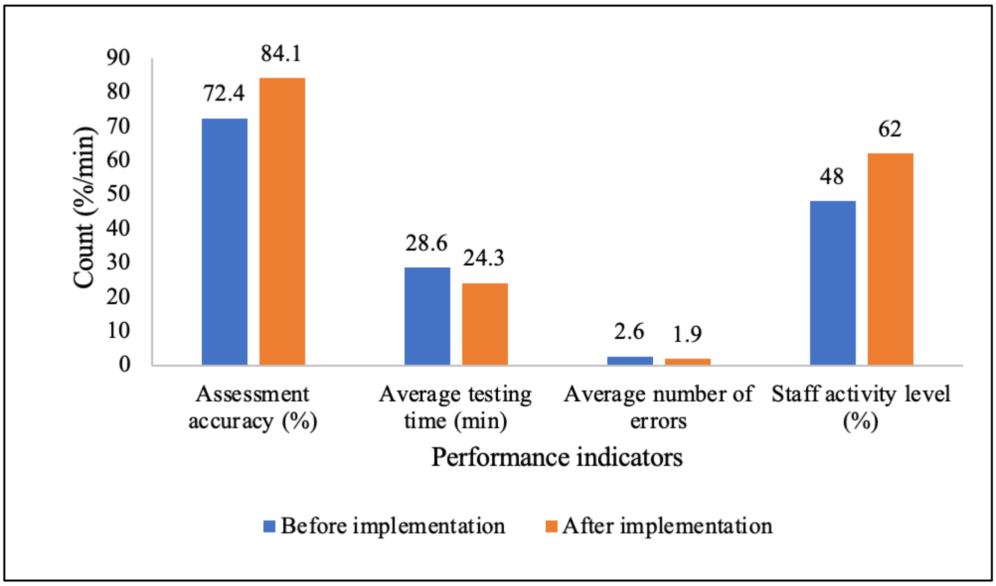

Figure 2: Comparison of Key Performance Indicators Before and After System Implementation

As shown in Figure 2, the proposed system leads to a clear improvement i all key performance indicators after implementation.

Discussion

The experimental results demonstrate that the integration of latent trait estimation, temporal behavioural indicators and semantic processing leads to a measurable improvement in both accuracy and stability of knowledge assessment. Unlike traditional Item Response Theory models that rely exclusively on response correctness, the proposed hybrid framework incorporates behavioural dynamics and qualitative response features, allowing a more comprehensive estimation of the latent ability parameter. The inclusion of response-time patterns reduces the probability of inflated ability scores caused by guessing or inconsistent answering behaviour, thereby enhancing measurement reliability.

From a theoretical perspective, the proposed model contributes to the development of multidimensional adaptive assessment systems by combining probabilistic modeling with behavioral and semantic analytics within a unified computational structure. This integration supports a richer representation of user competence and aligns with recent trends in hybrid adaptive assessment research, where behavioral data are increasingly considered as an essential component of knowledge evaluation.

From a practical standpoint, the system not only improves predictive accuracy but also optimizes the testing process by dynamically adjusting task difficulty and reducing redundant items. Such adaptability is particularly relevant in corporate training environments, where assessment efficiency and employee engagement are critical. The semantic evaluation of open-ended responses further enhances construct validity by incorporating qualitative evidence of understanding into the overall competence score (22).

However, the hybrid architecture increases computational complexity and requires calibrated semantic feature extraction procedures. Future research may focus on optimizing model scalability, refining behavioral parameter estimation and expanding validation across diverse educational and professional contexts.

Compared with existing adaptive testing systems, the proposed framework offers a broader integration of complementary data sources. Classical computerized adaptive testing (CAT) platforms typically rely on IRT-based item selection without incorporating response-time or semantic features (15, 19). Recent AI-driven automated scoring systems, such as those applying transformer-based language models for essay evaluation, address qualitative response analysis in isolation without connecting it to a latent trait model (23). Multimodal assessment frameworks that combine behavioral and semantic signals have been proposed in the educational data mining literature, but these rarely operate within a unified probabilistic structure that also supports real-time adaptive item selection (12, 21). The present model distinguishes itself by embedding all three components within a single coherent estimation pipeline, enabling simultaneous adaptation of item difficulty and competence estimation based on accuracy, timing and response quality.

The use of learner interaction data and behavioural analytics raises important considerations regarding privacy protection and data security. The proposed system processes response-time logs, interaction traces and open-ended textual responses, all of which may be considered personally identifiable information in certain regulatory contexts. In line with established ethical principles for learning analytics (14), the implementation of the system must include data anonymization procedures, role-based access control and transparent data governance policies. The security subsystem described in the architecture section provides authentication and audit trail mechanisms; however, institutional deployment should additionally comply with applicable data protection regulations and obtain informed consent from all participants. Future development of the platform should incorporate privacy-by-design principles to ensure that individual competence profiles are stored and transmitted securely.

The quantitative outcomes presented in Table 1 merit closer examination in relation to prior benchmarks in adaptive testing research. The 11.7% improvement in assessment accuracy observed in this study is consistent with previous findings (16), who demonstrated that IRT-based adaptive item selection consistently outperforms conventional fixed-format tests in terms of measurement precision. The reduction in testing time by 4.3 minutes aligns with the efficiency gains documented in large-scale computerized adaptive testing applications (17, 18, 24), where adaptive item selection reduces redundant items while maintaining comparable measurement accuracy. As illustrated in Figure 2, all four performance indicators show consistent positive change following system implementation, with the most pronounced effect observed in assessment accuracy and staff activity level. The 14% increase in staff activity level corroborates observations from learning analytics research (21, 25), indicating that personalized assessment trajectories contribute to higher learner engagement and motivation. The modular architecture illustrated in Figure 1 demonstrates how parallel processing of heterogeneous data streams enables simultaneous improvements across multiple evaluation dimensions. The IRT-based probabilistic module, the temporal modelling component and the semantic analysis module operate concurrently, each contributing complementary information to the final competence estimate. Taken together, these empirical outcomes confirm that the proposed hybrid integration strategy delivers measurable improvements over single-modality assessment approaches, supporting both the theoretical rationale and the practical utility of the framework.

Conclusion

This study presented an adaptive hybrid model for automated knowledge assessment integrating latent estimation based on Item Response Theory, temporal response dynamics and semantic analysis of open-ended answers within a unified computational framework. The developed architecture enables real-time adaptive testing, multimodal data processing and continuous updating of individual competence profiles.

The experimental validation confirmed that the proposed approach enhances the accuracy and reliability of knowledge level estimation while improving assessment efficiency in corporate environments. By combining probabilistic modeling, behavioral analytics and semantic evaluation, the system provides a more comprehensive representation of personnel competence compared to conventional non-adaptive testing methods.

The key findings of this study are summarized in Table 1, which shows comparative results of the before-and-after experimental evaluation across four performance dimensions: assessment accuracy, average testing time, average number of errors and staff activity level. All four indicators demonstrate consistent improvement following system implementation, confirming the overall effectiveness of the proposed framework. The architecture of the system, depicted in Figure 1, illustrates the modular design that enables integration of heterogeneous data streams within a unified adaptive pipeline. Figure 2 provides a visual comparison of the key performance indicators before and after implementation, further confirming the positive impact of the system across all measured dimensions. These results collectively demonstrate that the proposed hybrid model is both technically robust and practically effective in real-world deployment conditions.

Although the proposed model demonstrates promising results, several limitations must be acknowledged. First, the experimental validation was conducted within a single corporate IT environment with a relatively small participant group, which limits the generalizability of the findings. The sample size and professional homogeneity of the participants may not represent the full diversity of learners in educational or multi-domain assessment contexts. Second, the weighting coefficients w_1, w_2 and w_3 in the hybrid integration function were assigned based on expert judgement and domain-specific calibration rather than derived through systematic optimization. The sensitivity of the final competence estimate to variations in these weights has not been formally analyzed and different weight configurations may yield substantially different results. Third, the semantic analysis module relies on pre-trained language model embeddings, which may not adequately capture domain-specific terminology or non-standard phrasing common in specialized professional fields. The quality of the semantic similarity score depends heavily on the alignment between the training corpus of the embedding model and the target assessment domain. Fourth, the current implementation does not include a formal mechanism for detecting and mitigating response-style biases, such as systematic over-speeding or strategic guessing, which can distort both temporal and accuracy-based estimates. Addressing these limitations in future work will be essential to strengthen the validity and scalability of the proposed framework across diverse assessment settings.

The proposed framework demonstrates strong practical potential for corporate training, certification and professional development systems requiring precise and data-driven evaluation mechanisms. Future research will focus on expanding the model with additional data modalities, refining semantic assessment techniques and validating the approach across larger and more diverse datasets to ensure scalability and broader applicability.

From a broader perspective, the findings of this study support the growing consensus in educational measurement research that single-modality assessment approaches are insufficient for capturing the full complexity of human knowledge and competence. The convergence of psychometric theory, temporal modelling and natural language processing creates new opportunities for designing assessment systems that are simultaneously rigorous, adaptive and sensitive to the qualitative dimensions of learning. The proposed hybrid framework represents a step in this direction, providing a foundation for future work on multimodal competence evaluation. In particular, the integration of knowledge tracing models, advanced neural language representtations and learning analytics frameworks into the hybrid pipeline offers promising directions for enhancing both the theoretical depth and practical impact of automated knowledge assessment. As digital assessment environments continue to evolve, approaches that combine interpretability with computational efficiency will be essential for ensuring fair, accurate and meaningful evaluation outcomes across educational and professional domains.

Abbreviations

IRT: Item Response Theory, NLP: Natural Language Processing, ML: Machine Learning.

Acknowledgment

The author expresses sincere gratitude to the management and staff of the participating corporate organization for providing access to the testing environment and supporting the experimental implementation of the proposed assessment system.

Author Contributions

All authors were responsible for conceptualization, methodology development, system design, software implementation, data analysis, experimental validation and manuscript preparation.

Conflict of Interest

The authors declare that they have no conflict of interest regarding the publication of this paper.

Data Availability

The datasets generated and analyzed during the current study are not publicly available due to corporate confidentiality restrictions but are available from the author upon reasonable request.

Declaration of Artificial Intelligence (AI) Assistance

The corresponding author used an AI-based language assistance tool to improve grammar, clarity and academic style of the manuscript. The scientific content, methodology, analysis, results and conclusions presented in this work are entirely the original work of the authors.

Ethics Approval

The study was conducted in accordance with institutional guidelines. All participants were informed about the purpose of the assessment procedure and participation was voluntary. No personal sensitive data were disclosed or used outside the research framework.

Funding

This research received no external funding and was conducted as part of independent academic work.

References

1. Lord FM. Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associate.1980. doi: 10.4324/9780203056615
2. Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publications.1991. https://uk.sagepub.com/en-gb/eur/book/fundamentals-item-response-theory doi: 10.2307/2075521
3. van der Linden WJ. A hierarchical framework for modeling speed and accuracy on test items. Psychometrika. 2007;72(3):287–308. doi: 10.1007/s11336-006-1478-z
4. van der Linden WJ. A lognormal model for response times on test items. J Educ Behav Stat. 2006;31(2): 181–204. doi: 10.3102/10769986031002181
5. Klein Entink RH, van der Linden WJ, Fox JP. A Box-Cox normal model for response times. Br J Math Stat Psychol. 2009;62(3):621–640. doi: 10.1348/000711008X374126
6. van der Linden WJ, Glas CAW. Computerized adaptive testing: Theory and practice. Dordrecht: Kluwer Academic Publishers; 2000. ISBN: 978-0-7923-6425-0
7. Chang HH. Psychometrics behind computerized adaptive testing. Psychometrika. 2015;80(1):1–20. doi: 10.1007/s11336-014-9401-5
8. Landauer TK, Foltz PW, Laham D. An introduction to latent semantic analysis. Discourse Process. 1998;25(2–3):259–284. doi: 10.1080/01638539809545028
9. Graesser AC, McNamara DS, Louwerse MM, Cai Z. Coh-Metrix: Analysis of text on cohesion and language. Behav Res Methods Instrum Comput. 2004;36(2):193–202. doi: 10.3758/BF03195564
10. Williamson DM, Xi X, Breyer FJ. A framework for evaluation and use of automated scoring. Educ Meas Issues Pract. 2012;31(1):2–13. doi: 10.1111/j.1745-3992.2011.00223.x
11. Chandrasekaran D, Mago V. Evolution of semantic similarity: A survey. ACM Comput Surv. 2021;54(2): 1–37. doi: 10.1145/3440755

12. Romero C, Ventura S. Educational data mining: A review of the state of the art. IEEE Trans Syst Man Cybern C. 2010;40(6):601–618. doi: 10.1109/TSMCC.2010.2053532
13. Brusilovsky P, Millán E. User models for adaptive hypermedia and adaptive educational systems. In: Brusilovsky P, Kobsa A, Nejdl W, editors. The Adaptive Web. Lecture Notes in Computer Science, Vol. 4321. Berlin: Springer; 2007. p. 3–53. doi: 10.1007/978-3-540-72079-9_1
14. Pardo A, Siemens G. Ethical and privacy principles for learning analytics. Br J Educ Technol. 2014;45(3): 438–450. doi: 10.1111/bjet.12152
15. van der Linden WJ, editor. Handbook of item response theory. Vol. 1: Models. Boca Raton, FL: CRC Press; 2016. doi: 10.1201/9781315374512
16. Wauters K, Desmet P, Van den Noortgate W. Adaptive item-based learning environments based on the item response theory. Comput Educ. 2010;54(2):576–587. doi: 10.1016/j.compedu.2009.08.003
17. Eggen T, Straetmans G. Computerized adaptive testing for classifying examinees into three categories. Educ Psychol Meas. 2000;60(5):713–734. doi:10.1177/001316400219708
18. Weiss DJ, Kingsbury GG. Application of computerized adaptive testing to educational problems. J Educ Meas. 1984;21(4):361–375. doi: 10.1111/j.1745-3984.1984.tb01040.x
19. Segall DO. Computerized adaptive testing. In: Kempf-Leonard K, editor. Encyclopedia of Social Measurement. Vol. 1. San Diego, CA: Academic Press; 2004. p. 429–438. https://www.sciencedirect.com/referencework/9780123693983/encyclopedia-of-social-measurement doi:10.1007/0-306-47531-6_3
20. van der Aalst WMP. Process mining: Data science in action. 2nd ed. Berlin: Springer; 2016. doi: 10.1007/978-3-662-49851-4
21. Romero C, Ventura S. Educational data mining and learning analytics: An updated survey. Wiley Interdiscip Rev Data Min Knowl Discov. 2020;10(3):e1355. doi: 10.1002/widm.1355
22. Zhai X, Krajcik J, Pellegrino JW. On the validity of machine learning-based next generation science assessments: A validity inferential network. J Sci Educ Technol. 2021;30(2):298–312. doi: 10.1007/s10956-020-09879-9
23. Rogers A, Kovaleva O, Rumshisky A. A primer in BERTology: What we know about how BERT works. Trans Assoc Comput Linguist. 2020;8:842–866. doi: 10.1162/tacl_a_00349
24. Baker RS, Inventado PS. Educational data mining and learning analytics. In: Larusson JA, White B, editors. Learning Analytics: From Research to Practice. New York: Springer; 2014. 61–75. doi: 10.1007/978-1-4614-3305-7_4
25. Corbett AT anderson JR. Knowledge tracing: Modeling the acquisition of procedural knowledge. User Model User-Adap Inter. 1994;4(4):253–278. doi: 10.1007/BF01099821

How to Cite: Medetova KM, Latipova NX, Marisheva LT, Ravilov SM. Adaptive Hybrid Model for Knowledge Assessment Integrating IRT Estimation, Temporal Dynamics and Semantic Analysis. Int Res J Multidiscip Scope. 2026;7(3):923-934. DOI: 10.47857/irjms.2026.v07i03.010127