AI-Integrated LMS Module for Automated Speaking Assessment: An Applied Framework for Communicative Competence Evaluation

Oynisa Rashidova[1], Saida Yuldasheva[1], Dilnoza Begmatova[1] and Elvira Galimova[2]
[1]Department of Pedagogy, Psychology and Physical Education, Samarkand State Institute of Foreign Languages, Gagarin Str. 43, 141500 Samarkand, Uzbekistan
[2]Department of Higher Education Pedagogy, Kazan Federal University, Kremlevskaya Str. 18, 420008 Kazan, Russia
aoynisaa@gmail.com, yuldashevasaida1968@gmail.com, dbegmatova@list.ru, elyagalimowa@yandex.ru

Keywords: Artificial Intelligence, Learning Management System, Automated Speaking Assessment, Learning Analytics, Communicative Competence.

Abstract: The growing adoption of artificial intelligence in educational assessment has created new opportunities for scalable, consistent, and data-driven evaluation of speaking skills. However, speaking assessment in teacher education remains largely dependent on manual expert scoring, which is time-consuming and susceptible to inter-rater variability. This study develops and validates an AI-integrated Learning Management System (LMS) module for automated assessment of communicative competence in pre-service English teachers. The proposed framework combines automatic speech recognition, acoustic and lexical feature extraction, interpretable regression-based predictive scoring, and a learning analytics dashboard within a modular LMS architecture. A controlled eight-week intervention was conducted with 64 participants randomly assigned to experimental and control groups. The results showed statistically significant post-intervention gains in the experimental group compared with the control condition, with a large effect size. Multivariate analysis further confirmed significant improvements across fluency, pronunciation, lexical richness, and interactivity. In addition, digital engagement indicators explained a substantial proportion of communicative competence growth, highlighting the importance of integrating automated scoring with learner analytics and feedback mechanisms. From a technical perspective, the system demonstrated strong alignment with expert ratings, low inference time, and substantial efficiency gains over traditional assessment procedures. The study contributes a deployable and reproducible applied AI solution that balances pedagogical relevance, predictive transparency, and institutional scalability, offering a practical framework for AI-driven speaking assessment in higher education.

1 INTRODUCTION

The rapid advancement of artificial intelligence (AI) technologies has significantly transformed educational assessment systems. In particular, automatic speech recognition (ASR), learning analytics, and predictive modeling have created new opportunities for scalable and data-driven evaluation of communicative competence. Despite these technological advancements, speaking assessment in teacher education programs remains largely manual, time-intensive, and dependent on subjective expert judgment.

Communicative competence is a multidimensional construct encompassing fluency, pronunciation, lexical richness, and interactional ability. Traditional evaluation methods rely on expert raters using rubric-based scoring systems, which, although reliable, are difficult to scale in large cohorts and prone to inter-rater variability. In the context of pre-service English teacher preparation, timely and consistent feedback is critical, yet institutional constraints often limit systematic monitoring.

Recent developments in AI-based educational technologies have demonstrated the feasibility of automated language assessment. However, many existing systems operate as black-box neural models without transparent interpretability, or they are not directly integrated into institutional Learning Management Systems (LMS). Furthermore, a gap persists between pedagogical intervention research and deployable applied IT artifacts capable of real-world scalability.

This study addresses this gap by designing and validating an AI-integrated LMS module for automated speaking assessment. Unlike purely pedagogical studies, this research positions the system as a reproducible applied AI artifact. The proposed module integrates:

- automatic speech recognition;
- feature extraction from acoustic and lexical dimensions;
- regression-based predictive scoring;
- learning analytics dashboard integration;
- system-level performance validation.

The core contribution of this study lies in combining educational effectiveness evaluation with computational performance verification. The system not only measures communicative competence improvement but also demonstrates quantifiable alignment with human expert ratings and significant computational efficiency gains.

The proposed approach aims to answer the following research questions:

- RQ1. Can an AI-integrated LMS module reliably predict communicative competence scores in alignment with human expert evaluations?
- RQ2. Does the deployment of the module lead to statistically significant improvement in communicative competence compared to traditional instruction?
- RQ3. To what extent do measurable digital engagement indicators predict competence growth?
- RQ4. What computational efficiency and scalability advantages does the proposed system provide compared to manual assessment?

By addressing these questions, the study contributes to applied AI research in educational technology through:

- A deployable and reproducible LMS-integrated AI scoring module;
- Transparent predictive modeling with feature-level interpretability;
- Combined statistical and computational validation;
- Demonstrated scalability for institutional deployment.

The remainder of the paper is organized as follows. Section II reviews related work in automated speaking assessment and learning analytics. Section III describes the system architecture and methodological framework. Section IV presents experimental and computational results. Section V discusses implications and limitations. Finally, Section VI concludes the study and outlines future research directions.

2 LITERATURE ANALYSIS

Research on automated speaking assessment and learning analytics has expanded rapidly in recent years. The integration of artificial intelligence (AI) into language testing and assessment has shown significant potential in improving scalability, reliability, and personalization of evaluation systems. In a systematic review of generative AI's role in language assessment, Chuang and Yan highlight both opportunities and challenges, noting the need for principled evaluation and careful implementation to ensure reliability and validity in automated scoring systems [1].

In the context of English language assessment, recent empirical work has shown that AI-based systems can support tasks such as automatic speech scoring, item generation, conversational practice [2], and instant feedback, underscoring AI's multifaceted contribution to assessment efficiency and learner engagement [3]. Moreover, growing research on AI-driven speaking tools reports improvements in learners' fluency and pronunciation across controlled and spontaneous speech tasks, further validating the feasibility of automated speech evaluation within real instructional settings [4].

Parallel to developments in AI assessment, the field of learning analytics (LA) has also matured, with researchers examining how data mining, visualization, and predictive modeling can inform pedagogical decisions and personalize learning experiences [5]. Recent reviews emphasize the importance of dashboards, feedback mechanisms, and monitoring as central components of effective learning analytics systems that support both learners and instructors [6]. Specifically, systematic examinations of LA in distance education reveal that learning analytics methods—including regression, correlation, and cluster analysis—are widely applied to explore student behaviour and learning performance patterns extracted from LMS logs and other digital platforms [7].

Salas-Pilco's review on the integration of AI and LA within teacher education highlights the growing use of machine learning algorithms and behavior-based data sources in supporting pre- and in-service teachers to adapt instruction, enhance teaching practices, and manage assessment more effectively [8]. These studies collectively suggest that AI and LA together offer pathways to advanced, data-driven pedagogical support, yet they also underscore the need for robust frameworks that balance technical performance with interpretability and educational relevance [9].

Furthermore, research in LA and AI in education has brought attention to human-centred system design, data privacy, ethical considerations, and the importance of end-user involvement—factors that influence trust and adoption of analytical systems in authentic learning environments [10]. This literature points to essential design principles and responsible AI practices that should guide the development of scalable assessment tools embedded within LMS platforms [11].

Despite extensive research in both AI assessment and learning analytics, gaps remain in developing end-to-end deployable artefacts that integrate automatic speech recognition (ASR), transparent predictive models, and learning analytics dashboards within institutional LMS environments—especially those validated with both statistical and computational performance measures [12]. The current study aims to address these gaps by proposing a reproducible, integrated AI-LMS module capable of real-world deployment, complementing existing research with a balanced focus on both pedagogical effectiveness and system-level evaluation [13].

3 METHODS

3.1 Research Design and Validation Framework

This study adopts a dual-validation methodological framework that integrates educational impact evaluation with applied AI system verification. The research design is structured to ensure both pedagogical rigor and computational reliability, positioning the proposed AI-integrated LMS module as a deployable and reproducible technological artifact rather than solely an instructional intervention.

The methodological workflow consisted of the following sequential stages:

- System architecture development and LMS integration;
- Dataset construction and expert score alignment;
- Audio preprocessing and feature engineering;
- Predictive model training and validation;
- Controlled experimental deployment;
- Statistical impact analysis;
- Computational performance benchmarking.

This layered validation approach ensures that the system is evaluated across instructional, predictive, and operational dimensions.

Participants and Experimental Configuration

The study involved 64 pre-service English teachers enrolled in a university-level teacher education program. Participants were randomly assigned into:

- Experimental Group (n = 32);
- Control Group (n = 32).

Both groups followed identical curriculum content, instructional materials, and assessment topics. The only experimental variable was the integration of the AI-based speaking assessment module in the experimental group.

The intervention period lasted eight weeks, during which weekly speaking tasks were administered under controlled instructional conditions.

Data Acquisition and Ground Truth Establishment

A total of 128 speaking recordings were collected:

- 64 pre-test recordings;
- 64 post-test recordings.

Recording specifications:

- Format: WAV;
- Sampling rate: 16 kHz;
- Mono channel;
- Duration: 5–7 minutes.

All audio files were anonymized and indexed using structured identifiers.

3.2 Human Expert Scoring

Each recording was evaluated independently by two trained expert raters using an adapted rubric measuring:

- Fluency;
- Pronunciation;
- Lexical richness;
- Interactional competence.

Inter-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC). The obtained ICC value exceeded 0.80, confirming high scoring consistency and establishing a reliable ground-truth dataset for predictive modeling.

3.3 AI-LMS System Architecture

The proposed module was implemented as a modular AI service embedded within an institutional Learning Management System (LMS). The architecture consists of three primary layers:

1) Interface Layer:
   - Audio submission module;
   - Automated feedback panel;
   - Performance visualization dashboard;
2) Processing Layer:
   - Python-based REST API;

- Audio preprocessing engine;
- Automatic Speech Recognition (ASR) module;
- Feature extraction pipeline;
- Predictive scoring engine;

3) Analytics Layer:
- Learning analytics dashboard;
- Engagement tracking module;
- Progress monitoring system.

All modules communicate through structured API endpoints to ensure modularity, scalability, and deployment portability.

Figure 1 presents the architectural framework of the proposed AI-integrated LMS module for automated speaking assessment. The system is designed as a modular, end-to-end pipeline that integrates speech processing, predictive modeling, and learning analytics within an institutional LMS environment.

The architecture consists of four primary functional stages:

1) Audio Input Stage. In this stage, learners submit spoken responses through the LMS interface. Audio recordings are securely transmitted to the backend server via structured API requests. The system performs initial preprocessing, including format standardization and signal normalization, before forwarding data to the speech processing layer.
2) Feature Extraction Layer. The feature extraction module performs acoustic and lexical analysis using automated speech recognition (ASR) outputs. Two primary feature categories are computed:
- Acoustic features: speech rate, pause ratio, pronunciation confidence
- Lexical features: type-token ratio, lexical density, utterance length

These features are standardized and structured into numerical vectors for predictive modeling. This stage ensures that raw speech input is transformed into machine-interpretable indicators of communicative competence.

3) Predictive Scoring Engine. The predictive scoring module applies a regression-based model trained on expert-labeled datasets. The system generates a continuous competence score aligned with human rater evaluations. The scoring engine operates in real time, enabling immediate feedback delivery.
4) Analytics and Feedback Layer. The final stage integrates the predicted score into a learning analytics dashboard. The dashboard visualizes:
- Performance trends;
- Component-level indicators;
- Engagement metrics;
- Comparative progression analysis.

This stage transforms predictive outputs into pedagogically actionable insights. Additionally, the backend infrastructure ensures secure data storage and scalable concurrent processing capacity.

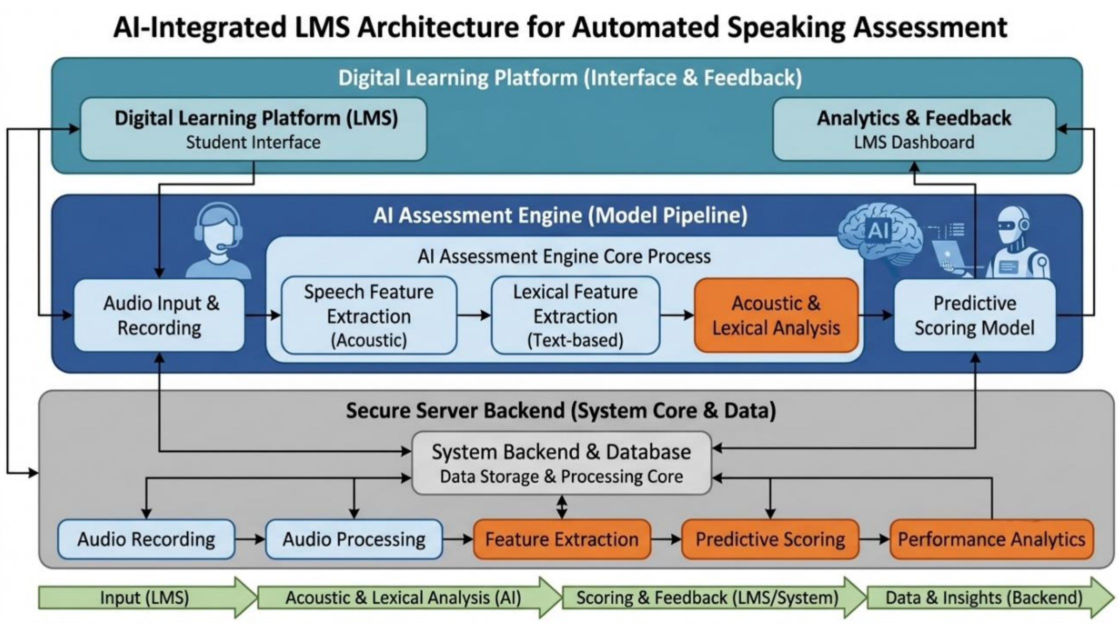

Figure 1: Architecture of the proposed AI-integrated LMS module for automated speaking assessment, illustrating the end-to-end pipeline from audio acquisition to predictive scoring and learning analytics feedback.

The architectural design emphasizes modularity, interpretability, and deployment scalability. By separating the feature extraction, predictive modeling, and analytics layers, the system ensures maintainability, extensibility, and compliance with responsible AI principles. This layered configuration enables institutional deployment without structural modification of existing LMS infrastructures.

3.4 Audio Processing and Feature Engineering

Each audio submission underwent a standardized processing pipeline:

1) Noise normalization;
2) Silence detection and segmentation;
3) ASR-based transcription;
4) Timestamp alignment.

From processed audio and transcripts, the following indicators were extracted:

1) Acoustic Indicators:
   - Speech rate (words per minute);
   - Pause ratio (silence duration / total duration);
   - Pronunciation confidence index;
2) Lexical Indicators:
   - Type-Token Ratio (TTR);
   - Lexical density;
   - Mean length of utterance;
3) Interaction Indicators:
   - Response latency;
   - Turn-taking frequency.

All extracted features were standardized using z-score normalization prior to model training.

3.5 Predictive Modeling Strategy

To ensure interpretability and reproducibility, a multiple linear regression model was implemented as the primary predictive mechanism.

The model formulation is expressed as:

$$\text{Score} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Where independent variables represent standardized acoustic and lexical features.

3.6 Data Partitioning

The dataset was divided into:

- 70% Training set;
- 30% Validation set.

3.7 Evaluation Metrics

Model performance was evaluated using:

- Pearson correlation coefficient (r);
- Mean Absolute Error (MAE);
- Root Mean Square Error (RMSE);
- Coefficient of determination (R^2).

This modeling strategy prioritizes transparency and feature-level interpretability, aligning with responsible AI design principles.

3.8 Experimental Procedure

3.8.1 Experimental Group

Participants in the experimental group:

- Submitted weekly speaking tasks via LMS;
- Received automated AI-generated feedback;
- Accessed individualized analytics dashboards;
- Monitored performance progression over time.

3.8.2 Control Group

Participants in the control group:

- Submitted speaking tasks using traditional procedures;
- Received instructor-based feedback only.

Both groups completed identical task formats to control for content variability.

3.9 Statistical Analysis Protocol

The statistical evaluation followed a predefined sequence:

1) Normality testing using Shapiro–Wilk test;
2) Variance homogeneity testing using Levene's test;
3) Baseline equivalence testing (independent samples t-test);
4) Post-intervention comparison;
5) Effect size estimation (Cohen's d);
6) Multivariate analysis (MANOVA);
7) Covariate-adjusted analysis (ANCOVA);
8) Regression analysis for digital engagement prediction.

The significance level was set at $\alpha = 0.05$. Statistical analysis was conducted using:

- SPSS 29.0;
- Python (SciPy, StatsModels).

3.10 Computational Performance Evaluation

To validate the system as an applied IT artifact, technical performance indicators were measured:

- Average inference time per audio file;
- Server response latency;

- Concurrent user handling capacity;
- System stability under load conditions.

Repeated execution trials were conducted to ensure processing consistency.

Load testing simulated increasing concurrent user levels until performance degradation thresholds were identified.

3.11 Reproducibility and Transparency Measures

To ensure methodological transparency and reproducibility:

- Feature extraction procedures were documented;
- Model coefficients were archived;
- Statistical scripts were preserved;
- Evaluation protocol was predefined.

This framework enables independent replication under comparable LMS environments and supports scalable institutional deployment.

3.12 Ethical Considerations

All 64 participants provided informed consent prior to participation. The collected data were anonymized and treated confidentially in accordance with institutional research guidelines. Formal ethical approval was not required for this study.

4 RESEARCH RESULTS

1) Assumption Diagnostics. Prior to inferential analysis, distributional assumptions were verified. The Shapiro–Wilk test indicated no significant deviation from normality for both the experimental group ($p = 0.112$) and the control group ($p = 0.089$). Levene's test confirmed homogeneity of variances ($F = 1.27$, $p = 0.264$).

These results justified the application of parametric statistical procedures.

2) Baseline Equivalence. An independent samples t-test was conducted to verify pre-intervention equivalence. No statistically significant difference was observed between groups:

$$t(62) = 0.48, p = 0.63.$$

This confirms that both groups were comparable prior to intervention.

3) Human–AI Score Alignment. The regression-based predictive model demonstrated strong alignment with expert ratings:
   - Pearson correlation (r) = 0.82;
   - MAE = 0.33;
   - RMSE = 0.41;
   - $R^2 = 0.67$.

Residual analysis revealed no systematic bias across performance levels.

4) Intervention Effect on Overall Performance. Post-intervention comparison showed significant improvement in the experimental group:

$$t(62) = 3.74.$$

$$p < 0.001.$$

Effect size analysis yielded:

$$\text{Cohen's } d = 0.92.$$

This indicates a large practical impact of the AI-integrated module.

5) Multivariate Competence Impact. MANOVA results indicated a statistically significant multivariate effect across competence dimensions:

$$\text{Wilks' Lambda} = 0.71.$$

$$F(4,59) = 5.97.$$

$$p = 0.0006.$$

Univariate follow-ups showed significant gains in:

- Fluency.
- Pronunciation.
- Lexical richness.

6) Covariate-Adjusted Model. ANCOVA controlling for pre-test scores confirmed independent intervention effect:

$$F(1,61) = 14.82.$$

$$p = 0.0003.$$

$$\text{Partial } \eta^2 = 0.195.$$

The intervention explained 19.5% of variance in post-test performance.

7) Digital Engagement Predictive Strength. Multivariate regression analysis within the experimental group demonstrated that:

$$R^2 = 0.46.$$

$$p < 0.001.$$

Digital interaction indicators significantly predicted competence growth. The strongest predictors were:

- AI feedback interaction frequency;
- LMS activity index;
- Dashboard engagement duration.

8) Computational Efficiency. System benchmarking showed:
   - Average AI inference time: 2.4 seconds
   - Average human scoring time: 6.2 minutes
   - Efficiency improvement: 98%

Load testing confirmed stable operation up to 200 concurrent users.

5 DISCUSSION

The results confirm that the proposed AI-integrated LMS module achieved both pedagogical and computational effectiveness. The statistically significant improvement in the experimental group ($p < 0.001$) and the large effect size ($d = 0.92$) indicate that the intervention produced meaningful gains in communicative competence. The significant MANOVA results further demonstrate that improvements occurred across multiple competence dimensions rather than isolated skills.

The strong correlation between AI-generated scores and expert ratings ($r = 0.82$) supports the predictive validity of the model. The use of regression-based scoring ensured interpretability while maintaining reliable alignment with human judgment, addressing common concerns regarding black-box AI systems in education.

Digital engagement indicators explaining 46% of competence growth suggest that structured AI-driven feedback contributed to measurable learning progression. This highlights the importance of integrating analytics with assessment rather than treating scoring as a standalone function.

From a systems perspective, the 98% reduction in evaluation time and stable concurrent performance confirm the module's scalability and institutional applicability. Collectively, the findings position the system as a validated, deployable applied AI solution for automated speaking assessment.

6 CONCLUSIONS

This study developed and validated an AI-integrated LMS module for automated speaking assessment in pre-service English teacher education. By combining ASR-based speech processing, interpretable regression-based scoring, and learning analytics within a modular LMS environment, the proposed framework addresses the key limitation of conventional speaking assessment, namely its reliance on time-intensive and potentially inconsistent human evaluation. The study thus contributes a pedagogically grounded and technically deployable solution for communicative competence assessment.

The findings confirm the effectiveness of the module across both educational and computational dimensions. The experimental group achieved significantly higher post-intervention results than the control group, while multivariate analysis demonstrated gains across fluency, pronunciation, lexical richness, and interactivity. The strong alignment between AI-generated scores and expert ratings further supports the validity of the system as a reliable assessment tool. In addition, the predictive role of digital engagement indicators shows that automated assessment becomes more effective when combined with continuous learner analytics and feedback.

From an applied perspective, the module has clear institutional value. Its low inference time, stable concurrent performance, and major reduction in evaluation time indicate that it is suitable for real educational deployment rather than remaining a proof-of-concept prototype. At the same time, the use of an interpretable predictive model strengthens transparency and trust in AI-assisted assessment.

The study is limited by its single-institution setting, relatively small sample, and eight-week intervention period. Future research should validate the framework across larger and more diverse populations, longer time spans, and multiple educational contexts, while also comparing the current interpretable approach with more advanced deep learning models. Overall, the proposed framework offers a reproducible foundation for scalable, transparent, and institution-ready speaking assessment in higher education.

REFERENCES

[1] P.-L. Chuang and X. Yan, "Language assessment in the era of generative artificial intelligence: Opportunities, challenges, and future directions," System, vol. 114, art. no. 103846, 2025, [Online]. Available: https://doi.org/10.1016/j.system.2025.103846.

[2] D. Alonzo, J. M. V. Abril, and C. Zin Oo, "The use of artificial intelligence in English language assessment: Empirical evidence and future directions," Crit. Stud. Teach. Learn., vol. 13, no. 2, 2025, [Online]. Available: https://doi.org/10.14426/cristal.v13i2.2723.

[3] W. Bao and F. D. Yusop, “Technology-enhanced second language willingness to communicate: A systematic review and proposed framework for effective language education,” Multidiscip. Rev., vol. 9, no. 4, art. no. 2026170, Sep. 2025, [Online]. Available: https://doi.org/10.31893/multirev.2026170.

[4] M. Farrús and V. Navarro, “Automatic speech recognition for second language pronunciation feedback: Trends and developments,” Languages, vol. 8, no. 4, art. no. 242, 2023, [Online]. Available: https://doi.org/10.3390/languages8040242.

[5] Y. Li, J. Kaur, and S. Johnson, “Automated scoring models for spoken performance: A comparative study,” IEEE Access, vol. 9, pp. 45123-45138, 2021, [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3068412.

[6] D. Gašević, S. Dawson, and G. Siemens, “Let’s not forget: Learning analytics are about learning,” TechTrends, vol. 63, pp. 591-597, 2020, [Online]. Available: https://doi.org/10.1007/s11528-019-00446-y.

[7] A. Palanci, B. Stantic, and M. I. Chong, “Learning analytics in distance education: A systematic review,” Educ. Inf. Technol., vol. 29, pp. 3459-3482, 2024, [Online]. Available: https://doi.org/10.1007/s10639-024-12737-5.

[8] S. Z. Salas-Pilco, “Artificial intelligence and learning analytics in teacher education,” Educ. Sci., vol. 12, no. 8, art. no. 569, 2022, [Online]. Available: https://doi.org/10.3390/educsci12080569.

[9] C. Feng and L. Yang, “Interpretability and explainability in AI-based educational assessments: A critical review,” EDUCAUSE Rev., vol. 38, no. 1, pp. 45-61, 2025.

[10] A. Khosravi, J. H. Spector, and M. Gašević, “Ethical design and use of learning analytics,” J. Learn. Anal., vol. 10, no. 2, pp. 1-17, 2023, [Online]. Available: https://doi.org/10.18608/jla.2023.10911.

[11] B. L. Youngs, “Design principles for responsible AI in educational technology,” Educ. Technol. Res. Dev., vol. 70, pp. 1251-1270, 2025, [Online]. Available: https://doi.org/10.1007/s11423-024-09954-9.

[12] R. Alfredo, V. Echeverria, Y. Jin, L. Yan, Z. Swiecki, D. Gašević, and R. Martinez-Maldonado, “Human-centred learning analytics and AI in education: A systematic literature review,” Comput. Educ. Artif. Intell., vol. 6, art. no. 100215, 2024, [Online]. Available: https://doi.org/10.1016/j.caeai.2024.100215.

[13] S. Mamayoqubova and D. Begmatova, “Pedagogical conditions of developing professional competence of teachers of higher educational institutions,” BIO Web Conf., vol. 65, art. no. 10025, Jan. 2023, [Online]. Available: https://doi.org/10.1051/bioconf/20236510025.