

Transforming Peer Assessment in Higher Education: An AI-Supported Approach Using the Peerscholar Platform

Svitlana Skvortsova¹, Tetiana Symonenko², Kira Hnezdilova³, Kateryna Niedialkova¹
and Nataliia Andrusiak⁴

¹*Department of Mathematics and Methods of its Teaching, State Institution "South Ukrainian National Pedagogical University named after K. Ushynsky", Staroportofrankovska Str. 26, 65029 Odesa, Ukraine*

²*Department of Ukrainian Philology and Social Communication, Bohdan Khmelnytsky National University of Cherkasy, Shevchenko Blvd. 81, 18031 Cherkasy, Ukraine*

³*Department of Primary and Special Education, Bohdan Khmelnytsky National University of Cherkasy, Shevchenko Blvd. 81, 18031 Cherkasy, Ukraine*

⁴*Department of Economics, Uman National University, Instytutska Str. 1, 20301 Uman, Ukraine
skvo08@i.ua, tetianasymonenko4@gmail.com, kiragnez@gmail.com, ndlvitaliy@ukr.net, andrusiak.n@gmail.com*

Keywords: Peer Assessment, Artificial Intelligence, Higher Education, Mathematics Teacher Training, Peerscholar, Methodological Competence.

Abstract: This study is devoted to a comprehensive analysis of peer assessment platforms with the support of artificial intelligence (AI) and empirical substantiation of the effectiveness of implementing the PeerScholar system in the process of professional training of future mathematics teachers. In the theoretical part of the work, a systematic comparison of nine leading intellectual platforms (RiPPLE, EduPCR, Peerceptiv, Kritik, PeerScholar, EvaluMate, Gradescope, Peergrade, ALEKS) was carried out, which allowed to identify the most relevant and economically affordable solutions for Ukrainian higher education institutions in conditions of limited resources. The empirical component of the study included 56 third-year students of the specialty "Secondary Education (Mathematics)", of which 53 respondents (n=53) participated in the final assessment, having completed five full cycles of peer evaluation of methodological developments in geometry during the semester. The use of a mixed-methods research design made it possible to combine quantitative data analysis of the questionnaire structure (32 points on the Likert scale, $\alpha=0.89$) with qualitative thematic analysis of open-ended responses and in-depth interviews. Statistical results demonstrated a high level of overall satisfaction with the platform (M=4.28, SD=0.62), with the methodological aspect of using the system receiving the highest priority in the perception of students (M=4.36, SD=0.61). It was found that future teachers showed a significantly higher interest in constructive feedback (M=4.81, SD=0.42) than in receiving grades (M=4.42, SD=0.52). A significant increase in inter-rater reliability of assessments from $r_s=0.51$ to $r_s=0.71$ over the semester was found, which indicates a progressive development of the assessment competence of future teachers. The study proves that the integration of AI frameworks not only optimizes the workload of the teacher, but also acts as a catalyst for deep methodological reflection, allowing students to effectively identify their own gaps in knowledge through the prism of critical analysis of the work of their classmates.

1 INTRODUCTION

Contemporary higher education is undergoing a fundamental transformation of assessment paradigms, in which traditional top-down evaluation is increasingly being replaced by participatory peer assessment models. Peer assessment is defined as an educational mechanism through which students evaluate the quality or learning outcomes of their

peers of equal status [1], [2]. This practice is critically important for the development of self-regulated learning skills, as it positions students in the role of critics and emerging experts [3], [4]. However, its implementation faces fundamental challenges, including score subjectivity, increased cognitive load for instructors, and the problem of "empty feedback," in which students rely on generic statements lacking methodological depth.

Historically, these limitations have constrained the scalability and trustworthiness of peer assessment as a tool for summative evaluation [1].

The emergence of artificial intelligence technologies-particularly natural language processing and machine learning algorithms - offers transformative potential to address these challenges [5], [6]. AI-enhanced platforms create comprehensive intelligent ecosystems that improve feedback quality by providing real-time analysis of draft comments to prompt specificity and actionable suggestions. They also aggregate and calibrate peer scores using sophisticated algorithms that estimate reviewer reliability based on historical performance and alignment with expert judgments [7] deliver meta-feedback on the quality of feedback itself to foster professional critique skills [5], [6] and support instructors through learning analytics that identify anomalous scoring discrepancies or emerging group conflicts. As a result, the instructor's role shifts from routine moderation to strategic learner support [1].

Ukrainian higher education institutions face unique challenges, including limited institutional budgets, outdated learning management systems, language barriers associated with predominantly English-language platforms, and the absence of comprehensive digital literacy support infrastructures [3]. Nevertheless, teacher education programs in mathematics require innovative approaches, as peer assessment offers dual benefits: the development of subject-matter expertise and assessment literacy essential for future pedagogical careers.

The aim of this study is to conduct a comparative analysis of AI-supported peer assessment platforms and to evaluate the effectiveness of PeerScholar in developing assessment competencies among pre-service mathematics teachers in Ukrainian higher education.

Research questions: RQ1: Which AI-enhanced peer assessment platforms provide access to free versions suitable for Ukrainian higher education institutions, and how do their functional features compare? RQ2: How do students in mathematics education programs perceive the impact of PeerScholar on the methodological, organizational, motivational, technological, and professional-reflective aspects of their learning? RQ3: What relationships exist between different dimensions of platform perception, and how do academic achievement levels moderate these perceptions?

2 EVOLUTION AND DIGITAL TRANSFORMATION OF PEER ASSESSMENT

Peer assessment has evolved through distinct phases of development, each characterized by increasingly sophisticated technological integration and pedagogical understanding [1], [4], [5], [7] with the modern phase from 2015 representing AI-enhanced peer assessment ecosystems that integrate natural language processing to analyze feedback quality, machine learning algorithms to model reviewer reliability, generative AI to structure feedback formulation, and learning analytics dashboards for instructor oversight, reflecting a fundamental rethinking from peer assessment as a substitute for assessment to a powerful learning mechanism.

Topping et al. [1] proposed a comprehensive framework of six process improvement areas, where reviewer assignment is optimized by algorithms based on trust metrics instead of random selection, individual review support is implemented through the EvalUMate platform's Eva chatbot, which analyzes draft comments and solves the problem of empty feedback [5] final scores are derived through Bayesian models that take into account reviewer bias [7], NLP-based feedback corpus analysis identifies patterns and misconceptions [8], [9], instructor support is implemented through anomaly detection panels, and integrated systems like RiPPLE coordinate all measurements to optimize results.

Empirical studies demonstrate pedagogical value through improved feedback quality, with AI-supported students providing significantly more specific and constructive feedback ($d=1.28$, $p<0.001$), development of metacognition through progressive improvement in the ability to distinguish between quality work [10], and scalability without compromising quality [1], but limitations have been identified in terms of over-reliance on AI suggestions, variable effectiveness across disciplines, and the requirement for initial investment in developing rubrics and calibrating algorithms. For prospective teachers, peer assessment provides authentic practice in formulating constructive feedback and recognizing common errors, which are critical pedagogical competencies [3].

Five groups of criteria were identified for the systematic analysis: accessibility (availability of a free plan, language support), assessment types (peer, self, 360), the role of AI (calibration, feedback

analysis), support for reflection, and integration with LMS. Particular attention was paid to the platforms' ability to support the methodological specificity of STEM disciplines, in particular mathematics.

Gradescope and PeerScholar stand out for their ability to work with handwritten works, which is critical for STEM disciplines. Peerceptiv uses Item Response Theory (IRT), and Kritik uses a "grading power" system to calibrate the accuracy of student assessments. ALEKS focuses on individual learning trajectories, while RiPPLE uses graph models for recommendations. EvaluMate uses LLM to provide assistance (scaffolding) to students while writing feedback in real time.

Access to tools varies significantly: free/research: RiPPLE (for academic research) and EvaluMate (for pilot projects) are the most accessible for experiments; 1) paid/institutional: Kritik imposes costs on students (\$29-49 per course), while Peerceptiv targets large institutions through institutional licenses.

For exact sciences, the following are best suited: Gradescope - for quick verification of massive computational tasks. ALEKS - for knowledge diagnostics and curriculum construction. Kritik and Peerceptiv - for ensuring high reliability of assessments on large streams of students. PeerScholar - work with photos of handwritten works, reflection phase.

Platform monetization models significantly affect accessibility for institutions with limited resources. Identified approaches include: "student pays" models (Kritik); institutional licensing (Peerceptiv, PeerScholar); Freemium models (Peergrade, Moodle Workshop); open source (Moodle Workshop); research/academic free projects (EvaluMate, EduPCR).

For Ukrainian HEIs, the critical factor is the cost of implementation. Platforms with a "student-pay" model (Kritik) can create barriers for low-income students. Systems with free individual plans (PeerScholar, Gradescope) or projects completely open for academic use (RiPPLE, EvaluMate) seem more acceptable.

PeerScholar is an online tool specifically designed to support the complex iterative process of peer evaluation. The platform offers three main modes (Phase Settings): 1. Classic Individual: standard mode, where each student works independently. 2. Case Study: a unique mode that allows the teacher to assign different cases to different students. Students can evaluate colleagues who worked on the same or different cases, which stimulates comparative analysis. 3. Group Work:

Support for team assignments where groups submit their work together and evaluate other groups. PeerScholar was chosen for our empirical study due to its unique three-phase architecture: Create - Assess - Reflect. Unlike systems that focus only on the final score, PeerScholar emphasizes the learning process through revision of both their own work and peer assessments, which creates unique opportunities for working with students of varying levels of proficiency. The tasks assigned to the students were primarily drawn from the school mathematics curriculum (i.e., they were not overly complex). The instructor's moderation of the entire process and their final assessment preclude the oversight of mathematical errors, methodological inaccuracies in comments, and other similar issues.

Student work algorithm:

- 1) Create phase. Upload work (methodological plans, handwritten solutions in various formats: text, images, videos and links to external resources).
- 2) Enable plagiarism detection (if necessary).
- 3) Assess phase. Receive several works of fellow groups.
- 4) Evaluate the work using interactive rubrics (Matrix, Point, Likert, Comment) and provide detailed feedback with an analysis of typical errors (if necessary - self-assessment).
- 5) Reflection phase (Reflect). Analysis of feedback on own work; evaluation of searches "feedback on feedback".
- 6) Review of other reviewers' grades (learning by comparison).

To minimize errors and inaccurate assessments, at the beginning of the experiment, students were provided with written instructions in Ukrainian on how to use the PeerScholar platform, followed by several trial (training) peer assessment sessions.

Algorithm of teacher's work:

- 1) Settings. Creation of rubrics, determination of number of reviewers.
- 2) Setting of schedule of automatic transition between phases.
- 3) Moderation. Analysis of dashboard "Student Progress & Grading", tracking of progress of each student, review of comments, removal of inadequate feedback, change of scores (if necessary).

3 EMPIRICAL RESEARCH METHODOLOGY

This study is based on the testing of the PeerScholar platform in the educational process of a pedagogical university in order to assess its impact on the methodological preparation of students. The testing was conducted during the 2024/2025 academic year within the framework of the academic discipline "Mathematics Teaching Methods". The participants were 56 third-year students of the specialty 014.04 Secondary Education (Mathematics).

The sample was characterized by the following indicators:

- 1) Average score (GPA) in mathematical disciplines: <70 (9%), 70-79 (28%), 80-89 (41%), 90-100 (22%).
- 2) Previous peer assessment experience: 38% had limited experience, 62% had no experience at all.
- 3) Frequency of use of digital technologies: constantly (28%), often (47%), sometimes (19%), rarely (6%).

Students completed five complex methodological tasks from the planimetry course:

- 1) Development of a lesson fragment for the Pythagorean theorem.
- 2) Analysis of common student errors in geometric proofs.
- 3) Methodological justification of teaching triangle similarity.
- 4) Comparative analysis of different approaches to introducing coordinate geometry.
- 5) Development of differentiated learning strategies for spatial geometry.

The effectiveness of PeerScholar was assessed through a mixed design. Quantitative method: questionnaire with 32 questions on a Likert scale (1-5) and 7 open-ended questions on six aspects (methodological, organizational, motivational, technological, professional-reflective, overall assessment). Qualitative method: interviews with 6 students of different levels of success, analysis of feedback and revisions of works. Objective indicators were also analyzed: time of completion of stages, number of revisions, inter-rater reliability.

Statistical processing in IBM SPSS Statistics 27 included descriptive statistics, correlation analysis (Spearman's ρ) and consistency assessment (Cronbach's α). Qualitative data were processed by thematic analysis with identification of categories and calculation of frequency of mentions.

4 QUANTITATIVE RESULTS OF THE APPROBATION

Descriptive analysis of the survey results (n=53, response rate 94.6%) showed a generally high positive perception of the students' experience of working with the PeerScholar platform. The overall satisfaction index was $M=4.28$ ($SD=0.62$), which corresponds to the assessment of "rather agree" with a tendency to "completely agree". The internal consistency of the questionnaire was high ($\alpha=0.89$), which confirms the reliability of the tools used.

Analysis of individual aspects revealed a differentiated perception of different characteristics of the platform. The students rated the methodological aspect of using PeerScholar the highest ($M=4.36$, $SD=0.61$), in particular: helped to realize their own methodological errors through the analysis of the work of their classmates ($M=4.51$, $SD=0.63$), and also improved the ability to analyze the methodological correctness of solutions ($M=4.38$, $SD=0.69$). Such data indicate a positive perception of the methodological aspects of the platform by students: they unanimously admit that the analysis of colleagues' work helps them better identify their own errors and critically assess the correctness of solving problems; the most significant result is that students showed a higher interest in constructive feedback ($M=4.81$, $SD=0.42$) than in receiving points ($M=4.42$, $SD=0.52$). Therefore, the data obtained confirm the unanimity of students in requesting meaningful professional communication, which is critically important for their professional development. The indicator in the question of "deeper understanding of the methodology for solving mathematical problems" ($M=4.21$, $SD=0.92$) indicates a certain differentiation: for some students, automated peer review became innovative, while others adapted to it more difficult. Rather, this may indicate the need for additional instruction before starting work with PeerScholar.

Calculation of the Spearman ρ correlation coefficient allowed us to identify the relationship between the following components of the methodological aspect of using PeerScholar: between the student's identification of alternative approaches to solving problems that had not been considered by him before and the improvement of the ability to analyze the methodological correctness of problem solutions ($r_s = 0.721$, $p<0.05$): the identification by students of new ways of solving problems during peer evaluation is a factor in increasing their ability to comprehensively methodically analyze solutions; between the

student's discovery of alternative approaches to solving problems that were not previously considered by him and the methodological value of formative (verbal) assessment ($r_s = 0.913$, $p < 0.01$): therefore, the verbal form of assessment becomes a key tool for the student to understand new experience; between the student's discovery of alternative approaches to solving problems that were not previously considered by him and the methodological value of constructive (advice) ($r_s = 0.697$, $p < 0.05$): the encounter with methodological novelty actualizes the students' need for meaningful interpretation and advice; between the improvement of the student's ability to analyze the methodological correctness of problem solutions and the methodological value of formative (verbal) assessment ($r_s = 0.714$, $p < 0.05$): the development of skills in methodological analysis of classmates' work leads to a higher professional assessment of the verbal form of feedback as the most informative; between the methodological value of formative (verbal) assessment and the methodological value of constructive (advice) ($r_s = 0.764$, $p < 0.05$), which indicate the convergence of different forms of qualitative assessment.

The above results prove that the use of PeerScholar stimulates the transition from "ordinary" problem solving to deep methodological reflection by the student - future teacher.

The professional-reflective aspect also received high marks ($M=4.35$, $SD=0.65$). Students noted the development of methodological skills ($M=4.74$, $SD=0.68$), communicative skills of feedback formulation ($M=4.81$, $SD=0.63$) and reflective skills ($M=4.54$, $SD=0.71$). It is important that 87% of students agreed or strongly agreed with the statement about the feasibility of using automated peer review in future school practice ($M=4.19$, $SD=0.63$).

The obtained values of Spearman's ρ (0.645-0.745 at $p < 0.05$) indicate statistically significant relationships between the key components of the professional-reflective aspect of using PeerScholar: between the development of a student's methodological skills and the development of his communicative skills ($r_s = 0.745$, $p < 0.05$): expert assessment of the work of classmates integrates analytical activity with the need to formulate reasoned and constructive feedback; between the development of a student's methodological skills and the development of his reflective skills ($r_s = 0.713$, $p < 0.05$): the experience of analyzing other people's solutions contributes to the extrapolation of objective assessment criteria to one's own activities,

which stimulates critical self-reflection; between the development of a student's methodological skills and the usefulness of reflection for learning ($r_s = 0.645$, $p < 0.05$): methodically prepared students tend to perceive reflection as a tool for professional improvement, rather than as a formal procedure; between the appropriateness of using automated peer review (like PeerScholar) in school mathematics practice and the development of reflective skills ($r_s = 0.679$, $p < 0.05$): the ability to critically analyze one's own mistakes correlates with the awareness of the appropriateness of implementing automated peer review in the future professional activities of a mathematics teacher.

The organizational aspect was characterized by moderate ratings ($M=3.78$, $SD=0.84$). Students found the structure of the work understandable ($M=4.40$, $SD=0.69$), but the estimates of the sufficiency of the time frame were lower ($M=3.40$, $SD=0.52$), which indicates the need to adjust the terms at the planning stage. Interestingly, students did not agree with the statement that PeerScholar was less time-consuming compared to traditional assessment ($M=2.89$, $SD=0.63$), but they recognized that these costs were justified by the learning outcomes ($M=4.31$, $SD=0.68$), and the algorithm for automatic distribution of work between students worked correctly ($M=4.61$, $SD=0.52$). The motivational aspect was characterized by moderately high ratings. Students confirmed that participation in automated peer review increased their responsibility for the quality of their own work ($M=4.33$, $SD=0.69$) and learning motivation ($M=3.97$, $SD=0.84$). Receiving feedback from multiple classmates was rated as more valuable compared to feedback from the instructor alone ($M=4.24$, $SD=0.75$). The possibility of anonymous evaluation had a positive effect on reducing students' anxiety ($M=4.16$, $SD=0.88$), although the level of confidence during the evaluation of works was moderate ($M=3.32$, $SD=0.94$), indicating the need for additional training of students for the role of reviewers.

Calculation of the Spearman's ρ correlation coefficient revealed a relationship between increased learning motivation and the value of multi-channel feedback from classmates in PeerScholar ($r_s = 0.707$, $p < 0.05$). This suggests that future teachers' orientation towards feedback from various sources contributes to the growth of their cognitive motivation.

The technological aspect received satisfactory ratings ($M=3.91$, $SD=0.79$), with significant variability in the perception of interface usability

($M=3.85$, $SD=1.02$). Correlation analysis showed a strong relationship between the assessment of the technical usability of the platform and overall satisfaction with the experience ($r_s = 0.68$, $p < 0.001$), which highlights the critical importance of the technical aspect for the overall perception. Most students (79%) had technically accessible devices for work, but 21% noted some difficulties with access, especially during distance learning.

Comparative analysis by achievement criterion revealed that students with higher average scores (≥ 85 , $n=22$) demonstrated statistically significantly higher ratings of the platform's methodological value ($M=4.61$ vs $M=4.18$) and professional development ($M=4.52$ vs $M=4.12$) compared to students with lower average scores (< 85 , $n=31$). However, the difference in overall satisfaction did not reach statistical significance ($M=4.38$ vs $M=4.19$), indicating the universal suitability of the platform for students with different levels of preparation. Thematic analysis of open-ended responses and interviews revealed five main categories of advantages of using PeerScholar in the professional training of future mathematics teachers: development of critical and methodological thinking (83%) - students identified their own mistakes through the analysis of other people's work, expanded their methodological tools; objectivity of assessment (65%) - three independent anonymous reviews reduce subjectivity; communication skills (58%) - formation of constructive feedback, important for working with students; flexibility of remote work (52%) - completion of tasks at a convenient time; reflection as a valuable stage (47%) - the opportunity to refine the work after comments. At the same time, the analysis also revealed a number of difficulties and limitations: time costs (71%), technical problems (38%), uncertainty in one's own assessment (34%), superficiality of peers' comments (29%). Additional analysis of data automatically collected by the PeerScholar platform provided objective information about students' behavioral patterns: the average assessment time was 47 min ($SD=18$), which exceeded the planned 30 min; 68% of students made revisions (average 3.4 corrections). Inter-rater reliability $r=0.62$ ($p < 0.001$) - moderate consistency, increasing from $r=0.51$ (first task) to $r=0.71$ (fifth), indicating the development of assessment skills.

5 DISCUSSION

The results of our study support the concept of "AI as an augments" rather than a replacement for the instructor. While AI effectively automates routine processes such as assignment distribution and score calibration, it is the human component - high-quality peer feedback and subsequent reflection - that enables deep learning. This finding aligns with the conclusions of Li [7] and Topping et al. [1], who argue that peer assessments can achieve expert-level accuracy when supported by intelligent calibration systems, as evidenced in our study by the increase in inter-rater reliability to $r = 0.71$.

Nevertheless, risks related to algorithmic bias and transparency must be acknowledged. Student interviews revealed a certain level of uncertainty regarding the correctness of their own assessments ($M = 3.32$, $SD = 0.94$), which corresponds with Guo's [5] findings on the potential reduction of student autonomy resulting from excessive reliance on AI-generated prompts. Consequently, the instructor's role shifts from direct evaluation of student work to the design of complex rubrics and the moderation of the 10-15% of cases in which the system detects assessment anomalies. The strong correlation identified between the perception of methodological novelty and the perceived value of textual feedback ($r = 0.913$, $p < 0.01$) highlights the social-constructivist nature of learning within PeerScholar. This finding aligns with the theory of the zone of proximal development, according to which feedback provided by peers operating at a similar cognitive level may be more comprehensible and acceptable to students than expert comments from instructors. At the same time, platform anonymity proved critical for reducing assessment-related anxiety ($M = 4.16$, $SD = 0.88$), although it must not result in irresponsibility - an issue mitigated through mechanisms that evaluate the quality of feedback itself. For mathematics-related disciplines, a key advantage of PeerScholar was its support for handwritten submissions. As noted by Nedialkova [3], mathematics cannot be fully transitioned to digital input without a loss of cognitive depth in problem-solving processes. Thus, the success of our pilot implementation can be attributed to the integration of traditional handwritten work with a digital moderation and feedback platform.

The primary limitation of this study is the absence of a control group, which prevents definitive attribution of improvements in assessment reliability solely to the platform rather than to overall course learning effects. Additionally, the English-language interface of PeerScholar remains a barrier for some students, reducing their task completion speed. It is worth noting that the PeerScholar environment exhibits limited flexibility for complex tasks.

Future research should therefore focus on longitudinal investigations of PeerScholar's impact on the actual pedagogical practices of future teachers in school settings, the use of generative AI to automatically create personalized rubrics tailored to specific mathematical and methodological tasks, and the analysis of how different reviewer assignment models influence learning quality, and the exploration of assessment methodologies for tasks requiring profound mathematical knowledge.

6 CONCLUSIONS

Overall, the study demonstrates that the integration of intelligent peer assessment platforms is a necessary condition for the modernization of higher pedagogical education in the context of digital transformation. A systematic comparative analysis revealed that the EdTech market offers differentiated solutions ranging from adaptive individualized learning systems to comprehensive reflective ecosystems, the selection of which should be guided by disciplinary specificity and institutional budget constraints. AI-supported peer assessment does not merely automate grading processes but creates a qualitatively new environment for professional dialogue, where errors function not as grounds for punishment but as resources for deep conceptual understanding and the development of pedagogical expertise.

The pilot implementation of PeerScholar confirmed its pedagogical and technological suitability for the preparation of future mathematics teachers. The integration of digital technologies, AI elements, and methodologically grounded peer assessment design fosters the development of modern assessment competence oriented toward reflection, objectivity, and professional responsibility. These findings align with Nediakova's [3] conclusions regarding the effectiveness of structured, multi-level peer assessment processes and support Guo's [5] observations on the critical role of reflection in the

development of students' metacognitive skills. Based on the pilot results, it is recommended to conduct an introductory training session on principles of constructive feedback and platform use prior to initiating peer assessment. This approach reduces initial student uncertainty and enhances feedback quality. Additionally, sufficient time should be allocated for each assessment stage, with a gradual increase in task complexity throughout the semester and selective quality control of student feedback accompanied by meta-comments on review quality. The empirical evaluation of PeerScholar in mathematics teacher education demonstrates that the digital transformation of assessment effectively addresses feedback scarcity. Statistical evidence confirms that students highly value the methodological benefits of peer review, perceiving it as a means of identifying their own cognitive gaps through the analysis of peers' alternative approaches. The observed increase in inter-rater reliability to $r = 0.71$ indicates the genuine development of assessment competence among future teachers within a single semester. These results justify recommendations for instructors to integrate peer assessment as a mandatory component of methodological training, preceded by explicit instruction in constructive critique principles; for university administrators to prioritize platforms with LTI integration and to establish institutional AI literacy policies; and for EdTech developers to focus on interface localization and the improvement of OCR modules for handwritten mathematical notation, thereby significantly lowering the technical entry barrier for STEM disciplines.

REFERENCES

- [1] K. J. Topping, E. F. Gehringer, H. Khosravi, S. Gudipati, K. Jadhav and S. Susarla, "Enhancing peer assessment with artificial intelligence," *International Journal of Educational Technology in Higher Education*, vol. 22, no. 1, art. 3, 2025, [Online]. Available: <https://doi.org/10.1186/s41239-024-00501-1>.
- [2] D. Babik, E. Gehringer, J. Kidd, K. Sunday, D. Tinapple and S. Gilbert, "A systematic review of educational online peer-review and assessment systems: Charting the landscape," *Educational Technology Research and Development*, vol. 72, no. 3, pp. 1653-1689, 2024, [Online]. Available: <https://doi.org/10.1007/s11423-024-10349-x>.
- [3] K. V. Nediakova, "Using artificial intelligence technologies for assessment and peer assessment of students," *Problems of Modern Transformations. Series: Pedagogy and Psychology*, vol. 9, pp. 1-12, 2025.

- [4] D. E. Pare and S. Joordens, "Peering into large lectures: Examining peer and expert mark agreement using peerScholar, an online peer assessment tool," *Journal of Computer Assisted Learning*, vol. 24, no. 6, pp. 526-540, 2008, [Online]. Available: <https://doi.org/10.1111/j.1365-2729.2008.00290.x>.
- [5] K. Guo, "EvaluMate: Using AI to support students' feedback provision in peer assessment for writing," *Assessing Writing*, vol. 61, art. 100864, 2024, [Online]. Available: <https://doi.org/10.1016/j.asw.2024.100864>.
- [6] C. Liu, J. Cui, R. Shang, Q. Jia, P. Rashud and E. Gehringer, "Generative AI for peer assessment helpfulness evaluation," in *Proceedings of the 17th International Conference on Educational Data Mining (EDM)*, 2024, pp. 412-419.
- [7] A. W. Li, "Using Peerceptiv to support AI-based online writing assessment across the disciplines," *Assessing Writing*, vol. 57, art. 100746, 2023, [Online]. Available: <https://doi.org/10.1016/j.asw.2023.100746>.
- [8] Wei et al., "Chain of thought prompting elicits reasoning in large language models," *arXiv preprint arXiv:2201.11903*, 2022.
- [9] W. Xiong, D. Litman and C. D. Schunn, "Natural language processing techniques for researching and improving peer feedback," *Journal of Writing Research*, vol. 4, no. 2, pp. 155-176, 2012, [Online]. Available: <https://doi.org/10.17239/jowr-2012.04.02.3>.
- [10] D. E. Paré, "The peerScholar process: Using formative peer and self-assessment to enhance and quantify critical thinking skills, metacognitive awareness, and sense of community," Ph.D. dissertation, University of Toronto, Toronto, Canada, 2020.