International Journal of Evaluation and Research in Education (IJERE)
Vol. 15, No. 6, December 2026, pp. 5105~5117
ISSN: 2252-8822, DOI: 10.11591/ijere.v15i6.40372

AI-supported formative assessment and STEM communication competencies in engineering education

Dilyara Dulatkyzy[1], Gulden Akbayeva[1], Kamalbek Berkimbayev[2], Murat Mirzakhmetov[3], Nuraim Ibragimova[4], Saniya Nurgaliyeva[5,6]
[1]Department of the Theory and Methodology of Foreign Language Training, Karaganda National Research University named after Academician Ye. A. Buketov, Karaganda, Kazakhstan
[2]Department of Pedagogy, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkestan, Kazakhstan
[3]Department of Philology, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
[4]Institute of Natural Science, School of Education, Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan
[5]Department of Primary Education, Faculty of Pedagogy and Psychology, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan
[6]Department of Russian Language and Literature, Ataturk University, Erzurum, Turkey

Article Info

Article history:

Received May 10, 2026
Revised Aug 6, 2026
Accepted Sep 22, 2026

Keywords:

AI-supported formative assessment
Collaboration
Critical thinking
STEM communication
Written communication

ABSTRACT

This study examined the association between an artificial intelligence (AI)-supported formative assessment intervention and students' science, technology, engineering, and mathematics (STEM) communication competencies, including written communication (WC), collaboration (COL), and critical thinking (CT), within a competency-based educational (CBE) context in Kazakhstan. A quasi-experimental pretest–posttest design was applied with 146 second-year engineering students. The intervention implemented a rule-based and natural language processing (NLP)-assisted formative feedback system, explicitly excluding generative AI models, to provide structured, rubric-aligned feedback on student performance. Results indicated statistically significant improvements in all competencies for both groups, with the experimental group (EG) demonstrating significantly greater gains and post-test differences between groups reaching statistical significance ($p<0.05$). Effect size analysis indicated moderate effects for WC ($d=0.50$), COL ($d=0.48$), and CT ($d=0.40$), while analysis of covariance (ANCOVA) results showed small-to-large intervention effects across competencies ($\eta^2p=0.08–0.15$). These findings suggest a positive association between AI-supported formative feedback and STEM communication competency development. The study contributes empirical evidence on the role of structured AI-assisted formative feedback in competency-based STEM education.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Saniya Nurgaliyeva
Department of Primary Education, Faculty of Pedagogy and Psychology
Abai Kazakh National Pedagogical University
Dostyk Ave., 13, Almaty, Kazakhstan
Email: sanianurgalieva77@gmail.com

1. INTRODUCTION

Large-scale engineering education increasingly faces a challenge in providing timely, individualized, and continuous formative feedback for students. As class sizes grow and competency-based educational (CBE) approaches require repeated performance evaluation, instructors often experience a "feedback bottleneck," where the demand for high-quality formative feedback exceeds available

instructional capacity. This issue is particularly relevant in science, technology, engineering, and mathematics (STEM) education, where communication competencies such as written communication (WC), collaboration (COL), and critical thinking (CT) require iterative practice and structured feedback for development. The rapid advancement of artificial intelligence (AI) has significantly transformed higher education by enabling adaptive learning environments, automated feedback systems, and data-driven instructional design [1], [2]. This transformation aligns with CBE, which prioritizes measurable learning outcomes over content coverage [3], and with formative assessment theory, which emphasizes continuous feedback for learning improvement [4]. These developments are increasingly relevant in STEM education, where such competencies are considered essential for employability in the context of Industry 4.0 [5].

STEM communication competencies are conceptualized as a multidimensional construct consisting of WC, COL, and CT as measurable indicators of communication performance in STEM contexts [6]. These competencies are defined as discipline-specific forms of scientific communication embedded in engineering problem-solving rather than general academic writing skills [7]. Unlike general language proficiency, STEM communication emphasizes argumentation, reasoning, and evidence-based explanation within technical domains. These competencies are increasingly emphasized within international engineering education and professional standards, where effective communication, teamwork, and analytical reasoning are considered essential attributes for workforce readiness and employability in contemporary engineering practice.

Empirical studies have shown that STEM-based learning environments can contribute to the development of transversal competencies, particularly communication and problem-solving skills [8], [9]. Similarly, research on AI in education indicates that intelligent tutoring systems, learning analytics, and automated feedback tools can enhance formative assessment and improve learning outcomes [10], [11]. However, evidence also suggests that these effects are highly dependent on instructional design quality and pedagogical integration, rather than the presence of AI tools alone [12]. Despite generally positive findings, several studies report negative or neutral outcomes when AI tools are implemented without structured pedagogy. These include reduced learner autonomy, superficial revisions, and limited transfer of skills [13]. Additional concerns involve over-reliance on automated feedback systems and lack of transparency in algorithmic decision-making, particularly in writing-related tasks [14]. These contradictions reflect an ongoing debate in the literature regarding AI as either an augmentation tool or a substitute for instructional processes [15]. Further limitations in existing research include the predominance of short-term interventions, small sample sizes, and insufficiently validated assessment instruments [16]. Moreover, most studies focus primarily on cognitive or technical outcomes, while systematic evidence on soft skills development within STEM communication remains fragmented [17]. This indicates a need for more robust, theory-driven, and competency-based evaluation frameworks.

In parallel, AI-supported formative feedback systems have been shown to improve learning efficiency by providing timely and individualized responses, particularly in writing tasks [18]. However, concerns remain regarding reduced learner autonomy, over-dependence on automated corrections, and limited transparency of scoring mechanisms [19]. From a sociocultural perspective, learning is mediated through structured interaction and scaffolding [20], [21]. Within this framework, the AI-supported intervention can be interpreted as a form of pedagogical scaffolding that supports learners during stages of performance that may not yet be independently achieved. Through timely and criterion-based feedback, the system may facilitate progression within learners' zone of proximal development (ZPD) by providing structured guidance during iterative revision processes rather than functioning as an autonomous replacement for instruction. Accordingly, this study proposes a conceptual framework in which AI-supported formative assessment operates as a mediating layer between instructional design (CBE-aligned tasks) and competency development (WC, COL, and CT), emphasizing iterative feedback loops within authentic STEM communication activities.

Despite growing interest in AI in education, several important gaps remain. First, although AI-based assessment systems have been widely investigated, existing research has focused predominantly on summative evaluation and cognitive performance outcomes, while evidence regarding the role of AI-supported formative feedback in developing STEM communication competencies remains limited [22]. In particular, there is insufficient evidence explaining how formative AI feedback influences CT and COL, which involve complex processes of reasoning, interaction, and knowledge co-construction beyond individual task performance. Second, there is limited evidence on how standardized competency frameworks, such as Association of American Colleges and Universities (AAC&U) VALUE rubrics, can be operationalized through AI systems in authentic classroom settings [23]. Third, integrated pedagogical models combining instruction, assessment, and AI-supported feedback within a unified competency-based framework remain relatively uncommon [24]. The novelty of this study lies in integrating competency-based instruction, formative assessment, and AI-supported feedback within a unified pedagogical framework for STEM communication development. Furthermore, the study simultaneously assesses WC, COL, and CT

using AAC&U VALUE rubrics, implements a hybrid rule-based and natural language processing (NLP)-assisted feedback mechanism that combines transparency with automated language analysis, and provides empirical validation in the underexplored higher education context of Kazakhstan.

Kazakhstan represents a transitional higher education context where CBE reforms and digital transformation initiatives are being implemented alongside traditional assessment practices [25], [26]. While national policies emphasize modernization of engineering education and alignment with international standards, AI-supported formative assessment remains weakly integrated into actual teaching practices. This creates an implementation gap between policy intentions and classroom-level assessment realities. The competency-based reform context may also provide favorable conditions for the intervention because competency-oriented learning emphasizes continuous performance monitoring, explicit learning criteria, and iterative feedback processes. Therefore, AI-supported formative assessment may align closely with current educational priorities by supporting structured competency development through timely and criterion-based feedback mechanisms. Against this background, Kazakhstan provides a relevant empirical setting for studying AI-supported formative assessment in a context of ongoing pedagogical reform. The aim of this study is to examine the association between an AI-supported formative assessment intervention and changes in students' STEM communication competencies (WC, COL, and CT) within a CBE framework. The study focuses on how structured AI feedback is integrated into instructional design rather than treated as a standalone tool.

The central research question guiding this study is: to what extent is AI-supported formative assessment associated with pre–post changes in students' STEM communication competencies (WC, COL, and CT) in a CBE setting in Kazakhstan? This study contributes to the literature by proposing an integrated conceptual model of AI-mediated formative assessment and by providing empirical evidence from an underexplored educational context. A particular contribution of the study is the implementation of a hybrid rule-based and NLP-assisted feedback approach as a pedagogical scaffold for STEM communication learning. Unlike purely automated scoring or generative AI systems, the proposed approach combines deterministic rule-based transparency with NLP-supported analysis to facilitate structured feedback processes aligned with competency development.

2. METHOD

2.1. Research design

This study employed a quasi-experimental nonequivalent groups pretest–posttest design. The design was selected because it enables the analysis of competency development over time in authentic educational settings where random assignment is not feasible due to institutional and ethical constraints in higher education contexts involving intact classroom groups. To strengthen internal validity, several control procedures were incorporated into the research design. Because participant allocation was based on existing classroom groups rather than random assignment, potential selection bias was addressed through baseline equivalence assessment using pretest measures of STEM communication competencies prior to intervention implementation. Pretest comparisons were conducted to examine whether significant differences existed between groups before treatment exposure. To ensure consistency of intervention implementation, both groups followed the same instructional content, learning objectives, task structure, and intervention duration. Standardized procedures for instructional delivery and assessment activities were applied across all sessions to minimize variation related to teaching conditions rather than intervention effects.

2.2. Participants

The study involved 146 second-year undergraduate students enrolled in the "Radio Engineering, Electronics and Telecommunications" program at Karaganda Buketov University. Participants were assigned to an experimental group (EG, n=78) and a control group (CG, n=68) based on intact classroom organization. The study was conducted in a Kazakh-medium STEM instructional context. Both groups followed the same curriculum, learning tasks, and were taught by the same instructor; however, only the EG received AI-supported formative feedback. A convenience sampling technique was employed because participants were selected from existing intact classroom groups that were accessible within the institutional setting. This approach was considered appropriate for the quasi-experimental design, where random sampling and random assignment were not feasible due to organizational and ethical constraints.

Participants were included if they were enrolled in the STEM communication in English (STEM-CE) course, completed the full intervention, and provided complete pre- and post-test data. From an initial sample of 154 students, 8 were excluded due to incomplete participation, resulting in a final sample of 146 students. The mean age was 19.8 years (SD=0.74). A priori statistical power analysis was conducted using G*Power 3.1 to assess whether the sample size was sufficient for detecting intervention effects [27]. Assuming a medium effect size (f=0.25), a significance level of α=0.05, and statistical power of 0.80,

the required minimum sample size was estimated to be approximately 128 participants. The final sample of 146 students therefore exceeded the minimum required sample size and was considered adequate for detecting meaningful intervention effects.

Background variables including English proficiency (Common European Framework of Reference for Languages (CEFR) B1–B2), cumulative grade point average (GPA), and prior experience with AI-based learning tools were collected to ensure baseline comparability, with no significant differences observed between groups. Participation was voluntary, and all students provided informed consent. Data were anonymized and processed in accordance with institutional ethical guidelines. The demographic characteristics of the participants are presented in Table 1.

Table 1. Demographic characteristics of participants

Variable	EG	CG	Total
Mean age (SD)	19.7 (0.72)	19.9 (0.75)	19.8 (0.74)
Male	48 (61.5%)	40 (58.8%)	88 (60.3%)
Female	30 (38.5%)	28 (41.2%)	58 (39.7%)
Mean GPA (SD)	3.21 (0.34)	3.18 (0.37)	3.20 (0.35)
English proficiency (CEFR)	B1–B2	B1–B2	B1–B2
Mean English placement score (SD)	71.4 (6.8)	70.9 (7.1)	71.2 (6.9)
Prior experience with AI-assisted learning tools	29 (37.2%)	24 (35.3%)	53 (36.3%)

2.3. Procedure

Data collection was conducted from February to April 2025 over an eight-week intervention period. At the beginning of the study, both the EG and CG completed pre-test tasks assessing baseline levels of WC, COL, and CT in STEM communication contexts. Following the pre-test phase, both groups participated in the same STEM communication instructional module incorporating case-based learning activities. The EG received AI-supported formative feedback integrated into the learning process, while the CG received conventional instructor-mediated feedback. At the end of the intervention, both groups completed post-test tasks that were equivalent to the pre-test in structure, complexity, and competency focus.

All tasks were evaluated using the same AAC&U VALUE rubric-based criteria applied consistently across pre- and post-assessments. Instruction in both groups was delivered by the same course instructor to ensure consistency of pedagogical implementation and to minimize instructor-related variability. To improve the systematic presentation of the study procedure, a research flowchart was added in Figure 1, illustrating pre-test administration, intervention implementation, post-test assessment, and data analysis stages. The flowchart provides a visual representation of the sequence of methodological procedures and the progression of participants throughout the study. To ensure consistency and fidelity of intervention implementation, monitoring procedures were conducted throughout the eight-week period. A standardized instructional plan, including identical learning objectives, learning materials, task structures, and implementation schedules, was followed for both groups. Periodic monitoring of instructional activities and assessment procedures was carried out to verify adherence to the intervention protocol and to reduce variability associated with implementation differences.

2.4. Intervention

The intervention was implemented within the STEM-CE course at Karaganda Buketov University, Kazakhstan. It employed an AI system designed to provide automated formative feedback on student writing tasks. The system was explicitly developed as a hybrid rule-based and NLP-assisted scoring engine rather than using generative large language models, ensuring interpretability, pedagogical transparency, and traceability of assessment decisions. To ensure alignment between instructional design and targeted competencies, Table 2 summarizes the main task types and their associated cognitive and communicative functions.

The AI system combined rule-based linguistic analysis with NLP-assisted feature extraction, including keyword and discourse-marker detection as well as argument-structure identification. The processing pipeline consisted of text preprocessing, linguistic feature extraction, rule-based detection of discourse and argumentation patterns, and alignment of extracted features with AAC&U VALUE rubric descriptors. The scoring model was calibrated using expert-validated samples to ensure consistency and interpretability.

The NLP process involved several sequential stages. During preprocessing, student texts underwent tokenization, stop-word removal, normalization, and sentence segmentation to improve feature extraction accuracy. The processed texts were subsequently analyzed using a hybrid algorithm combining term frequency–inverse document frequency (TF-IDF)-based lexical representation with deterministic rule-based pattern matching to identify discourse structures, logical transitions, and argumentation components. Feature extraction was operationalized across four dimensions: i) lexical cohesion, measured via TF-IDF similarity

scores; ii) discourse structure, identified through transition markers and logical connectors; iii) argumentation quality, operationalized through claim–evidence–reasoning structures detected via rule-based pattern matching; and iv) COL indicators, inferred from linguistic markers of peer referencing and interaction in written responses.

Feedback generation was implemented through a deterministic rule-based scoring procedure in which extracted features were mapped onto rubric levels using threshold-based classification rules. Scoring weights were defined through expert validation and refined during pilot testing to ensure alignment with competency descriptors. For each submission, the system generated criterion-specific formative feedback aimed at improving argument clarity, evidence integration, and logical coherence. Figure 2 illustrates the architecture and workflow of the AI-supported formative assessment system. Student submissions first undergo text preprocessing, including tokenization, normalization, and sentence segmentation. The processed text is subsequently analyzed through NLP-based feature extraction and classified using deterministic rule-based procedures before being aligned with AAC&U VALUE rubric descriptors. Based on the extracted features and rubric mapping, the system generates structured formative feedback that supports iterative student revision and resubmission cycles.

The AI system functioned exclusively as a formative support tool designed to facilitate iterative learning rather than summative evaluation. It complemented instructor feedback by providing immediate, structured, and rubric-aligned guidance aimed at improving argument clarity, evidence integration, and logical coherence during revision activities. To ensure transparency and reproducibility, all scoring rules were deterministic and fully traceable to observable linguistic features. System outputs were periodically reviewed by the course instructor and an independent assessment expert. Review procedures focused on rubric alignment accuracy, relevance of feedback, and interpretability of system outputs.

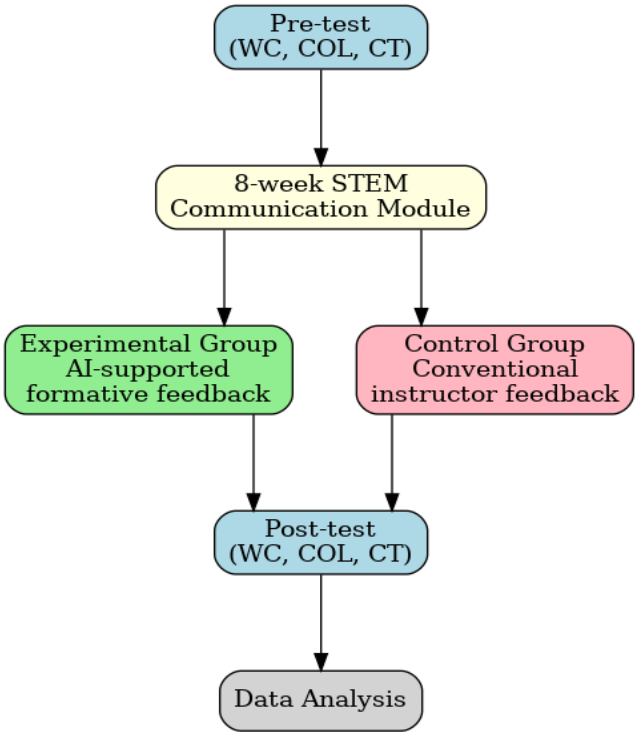


Figure 1. Research flowchart of the study procedure

Table 2. Types of learning activities and targeted competencies

Learning activity	Cognitive/communication action	Targeted competency
STEM case studies	Problem analysis	CT, WC
Analytical writing tasks	Structuring arguments	WC
Problem-solving communication	Explanation of solutions	WC, CT
Evidence-based writing	Integration of evidence	WC, CT
Revision tasks	Reflection and improvement	WC
Peer interaction	Collaborative meaning-making	COL

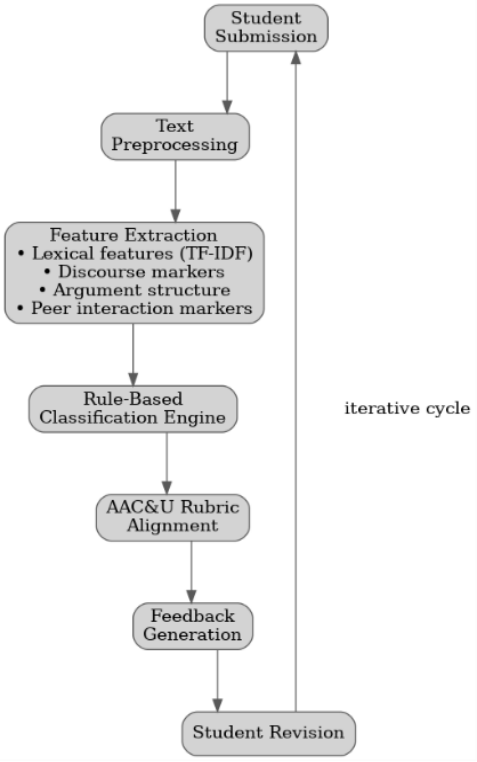


Figure 2. AI-supported feedback workflow

During pilot implementation, agreement between AI-based classifications and instructor evaluations demonstrated acceptable consistency (intraclass correlation coefficient (ICC)=0.84). In addition, random samples of feedback were manually inspected throughout the intervention to ensure stability and mitigate classification bias. An illustrative revision cycle demonstrated how students initially produced analytical texts with weak evidence integration or underdeveloped argument structures. The system generated targeted feedback prompts encouraging clarification of claims, strengthening of evidence use, and improvement of logical transitions. Students subsequently revised and resubmitted improved versions across iterative cycles. The system was deployed via a web-based learning platform integrated into the STEM-CE course. Students received individualized formative feedback immediately after submission, enabling timely revision within weekly instructional activities. To minimize potential confounding effects, both groups were taught by the same instructor, followed the same curriculum, completed identical learning tasks over the same eight-week period, used the same assessment rubric, and differed only in the provision of AI-supported formative feedback. Table 3 presents the structure of the intervention.

Table 3. Eight-week structure of the intervention

Week	STEM case focus	EG (AI-supported feedback progression)
1	Engineering problem analysis	Structured feedback (clarity+organization)
2	Technical explanation writing	Coherence-oriented feedback
3	Problem-solving argumentation	Argument quality feedback
4	Evidence-based writing	Evidence-use focused feedback
5	Collaborative STEM case	Integrated AI+peer interaction feedback
6	Analytical report writing	Multi-aspect feedback (WC+CT)
7	Integrated STEM task	Multi-criteria competency feedback (WC+CT+COL)
8	Final synthesis task	Holistic formative feedback (all competencies)

2.5. Instruments

The study employed adapted AAC&U VALUE rubrics developed by the AAC&U [28]. Each competency dimension was operationalized through analytically derived performance descriptors to ensure consistent evaluation across measurement points. The AAC&U VALUE rubrics were translated into Kazakh and conceptually aligned with the original English version to ensure equivalence of the assessment criteria in

a Kazakh-medium STEM communication context. The adaptation followed a multi-stage process, including initial translation into Kazakh by a bilingual specialist in academic communication, followed by independent back-translation into English by another bilingual expert to verify semantic equivalence. Minor linguistic refinements were introduced after comparison of the original and back-translated versions.

The rubric adaptation process additionally included contextual modification of selected descriptors to ensure relevance for STEM communication tasks in engineering education. Adaptation focused on preserving the conceptual meaning of the original AAC&U constructs while adjusting examples and task-related wording to fit the disciplinary context. These modifications were restricted to contextual language adaptation and did not alter the underlying competency structure of the original framework.

The adapted rubrics were validated by an expert panel consisting of specialists in STEM education, applied linguistics and academic writing, and educational assessment. The panel evaluated both linguistic and conceptual equivalence and confirmed alignment with the targeted competencies: WC, COL, and CT within STEM communication tasks. A pilot study involving 18 second-year engineering students enrolled in a comparable course was conducted to assess clarity, applicability, and inter-rater reliability. Students completed one STEM writing task that was independently evaluated by the course instructor and a trained assessor.

Inter-rater reliability demonstrated acceptable consistency (ICC=0.82), and minor refinements were subsequently introduced to improve descriptor clarity. To further establish construct validity and internal consistency, psychometric analyses were conducted prior to the main study. Exploratory factor analysis (EFA) was initially performed to examine the dimensional structure of the adapted rubric and verify the clustering of indicators across WC, COL, and CT domains. The results supported the expected three-factor structure. Confirmatory factor analysis (CFA) was subsequently conducted to assess model fit and demonstrated acceptable fit indices (comparative fit index (CFI)=0.93, Tucker–Lewis index (TLI)=0.91, root mean square error of approximation (RMSEA)=0.06), supporting construct validity of the adapted framework. Internal consistency analysis also indicated acceptable reliability across competency dimensions, with Cronbach's alpha coefficients ranging from 0.81 to 0.88 and an overall scale reliability of α=0.85.

Main performance data were collected using equivalent pre-test and post-test writing tasks aligned in complexity, format, and competency focus. The same rubric framework was applied at both stages to ensure longitudinal comparability. The AAC&U VALUE rubrics were used as competency-based assessment frameworks for evaluating competency development rather than as summative grading instruments. Final competency scoring was conducted by human evaluators using the adapted rubric criteria. An AI-supported formative feedback system was employed exclusively as a pedagogical intervention tool in the EG. The system generated formative feedback based on NLP techniques and AAC&U VALUE-aligned descriptors addressing organization, clarity, argumentation, evidence integration, critical reasoning, and COL-related indicators. The system was not used for competency scoring, grading, or statistical analysis.

Peer- and self-assessment procedures were conducted using the same adapted rubric framework to maintain consistency across evaluation sources. Final competency scores were calculated using a weighted aggregation model. Instructor ratings received the highest weighting due to their role as the primary reference standard, whereas peer- and self-assessments contributed equally as secondary evaluation sources. Weighting coefficients were determined a priori based on the instructional role and evaluative expertise of each assessment source. All evaluators received the same scoring guidelines and rubric descriptors prior to assessment to ensure scoring consistency across raters. To ensure comparability across raters and minimize scale-related variation, all scores were standardized (z-scores) prior to aggregation. The resulting composite scores reflected a balanced multi-source evaluation of each competency.

2.6. Data analysis

Data analysis included descriptive and inferential statistics to examine competency development across groups and measurement points. Descriptive statistics (means and standard deviations) were used to summarize competency scores, whereas inferential analyses examined within-group and between-group differences over time. Analysis of covariance (ANCOVA) was employed as the primary analytical approach to compare post-test scores between groups while controlling for baseline pre-test differences. This approach was considered appropriate given the quasi-experimental nonequivalent groups design. Prior to conducting ANCOVA, statistical assumptions were evaluated. Normality of residuals was assessed using the Shapiro–Wilk test, homogeneity of variances using Levene's test, and homogeneity of regression slopes through interaction analysis between group membership and pre-test scores. Results indicated no significant violations of assumptions (Shapiro–Wilk $p>0.05$; Levene's test $p=0.317$; interaction effect $p=0.421$). To improve interpretability, 95% confidence intervals (95% CI) and effect sizes were reported where appropriate alongside significance testing. All analyses were conducted using SPSS (v.26) with a significance level of $p<0.05$.

3. RESULTS

The descriptive analysis indicated improvements in all three competencies—WC, COL, and CT—in both the EG and CG from pre-test to post-test. Preliminary assumption testing indicated no significant violations of parametric test assumptions. Specifically, Shapiro–Wilk tests suggested that all variables were approximately normally distributed, and Levene's tests confirmed homogeneity of variances across groups ($p>0.05$ for all comparisons). The EG demonstrated greater gains across all competencies compared to the CG. The largest improvements were observed in WC and CT within the EG. Detailed descriptive statistics are presented in Table 4.

Table 4. Pre-test and post-test competency scores for EG and CG

Competency	Group	Pre-test mean (SD)	Post-test mean (SD)	Mean gain
WC	EG	2.10 (0.48)	2.55 (0.50)	+0.45
	CG	2.05 (0.46)	2.25 (0.47)	+0.20
COL	EG	2.20 (0.50)	2.55 (0.52)	+0.35
	CG	2.10 (0.49)	2.25 (0.50)	+0.15
CT	EG	2.30 (0.47)	2.75 (0.51)	+0.45
	CG	2.20 (0.48)	2.45 (0.49)	+0.25

Paired samples t-tests demonstrated statistically significant improvements in both groups across all competencies ($p<0.05$), with substantially greater gains observed in the EG. Post-test comparisons further revealed significant between-group differences favoring the EG across all competencies. Key inferential statistics are presented in Tables 5 and 6.

Table 5. Within-group changes (pre–post)

Competency	Group	t	df	*p*	Cohen's d
WC	EG	7.95	77	<0.001	0.90
	CG	3.30	67	<0.01	0.40
COL	EG	6.18	77	<0.001	0.70
	CG	2.48	67	<0.05	0.30
CT	EG	7.52	77	<0.001	0.85
	CG	4.13	67	<0.001	0.50

Table 6. Independent samples t-tests for post-test comparisons

Competency	t	df	*p*	Cohen's d
WC	3.01	144	<0.01	0.50
COL	2.88	144	<0.05	0.48
CT	2.41	144	<0.05	0.40

To control for baseline variability, one-way ANCOVA were conducted using post-test scores as dependent variables and pre-test scores as covariates. ANCOVA was additionally conducted to adjust for baseline differences between groups. The ANCOVA results indicated that adjusted post-test differences remained statistically significant in favor of the EG across all competencies. As shown in Table 7, the observed partial eta squared values indicate meaningful effects of the intervention after controlling for baseline scores, with WC ($\eta^2p=0.12$) reflecting a medium effect, COL ($\eta^2p=0.08$) a small-to-medium effect, and CT ($\eta^2p=0.15$) a large effect.

Table 7. ANCOVA results controlling for pre-test scores

Competency	F	df	*p*	Partial η^2
WC	19.50	(1,143)	<0.001	0.12
COL	12.43	(1,143)	<0.001	0.08
CT	25.24	(1,143)	<0.001	0.15

Figure 3 presents mean competency gains together with 95% confidence interval error bars. Across all competencies, the EG demonstrated consistently larger gains than the CG. The relatively limited overlap of confidence intervals further supports the observed differences between instructional conditions.

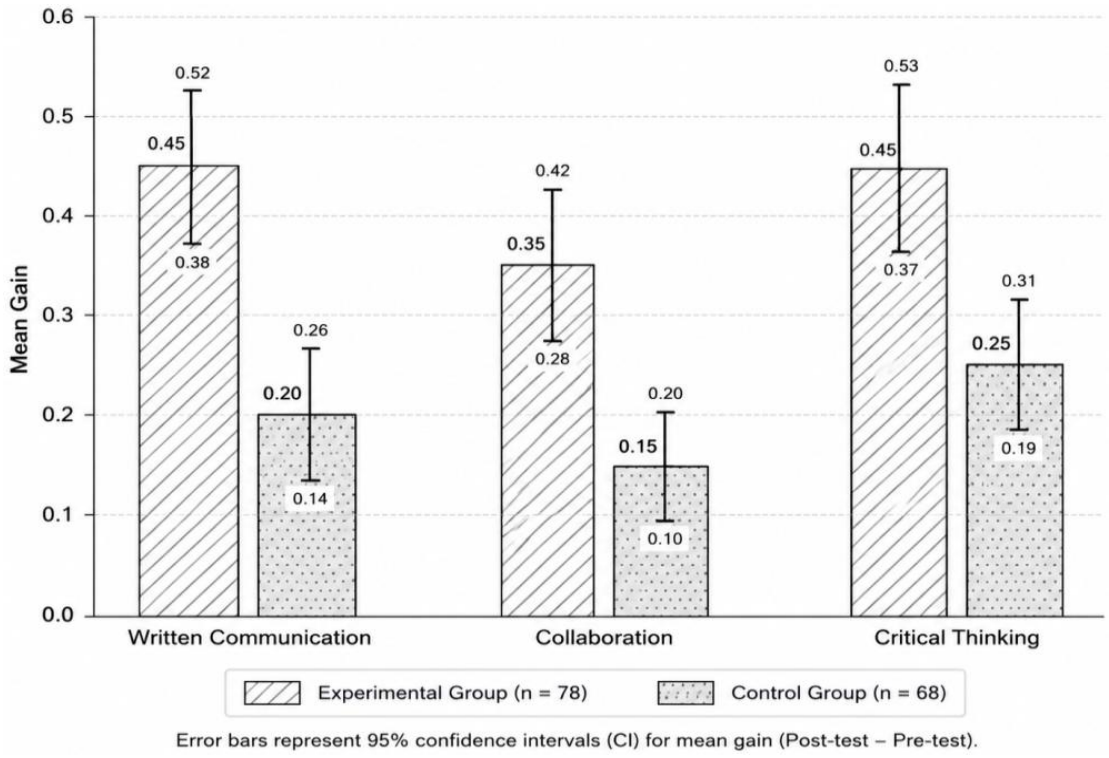


Figure 3. Pre-post competency gains by group (95% CI)

4. DISCUSSION

The findings suggest that AI-supported formative assessment was associated with stronger development of STEM communication competencies compared with conventional feedback practices. Although improvements were observed in both groups, the consistently greater gains demonstrated by the EG suggest that structured AI-supported feedback may have provided more frequent opportunities for iterative revision, self-monitoring, and criterion-based reflection during learning. By contrast, the comparatively smaller improvement in COL may reflect the more socially situated nature of collaborative competence, which depends on interpersonal interaction and shared regulation beyond individual text revision. This pattern is consistent with previous studies demonstrating that AI-supported formative assessment is particularly effective in improving higher-order communication skills, writing performance, and critical reasoning through iterative feedback processes [29].

From a CBE perspective, these findings may reflect the alignment between continuous feedback processes and measurable learning outcomes. CBE emphasizes continuous performance improvement through explicit competency indicators, and the AI-supported feedback system may have provided structured opportunities for repeated revision against defined criteria. The particularly strong gains observed in WC and CT may reflect the role of immediate feedback in supporting refinement of argument structure, evidence use, and analytical reasoning. This interpretation is consistent with CBE theory, which emphasizes continuous formative assessment, explicit learning outcomes, and iterative performance improvement as central mechanisms for competency development [30].

A possible explanatory mechanism concerns the degree of alignment between AI-generated feedback and the cognitive processes underlying each competency. WC and CT are primarily expressed through observable textual features such as argument structure, evidence integration, coherence, and logical progression. These features can be effectively addressed through NLP-assisted formative feedback. In contrast, COL involves socially situated behaviors, including negotiation, shared regulation, interpersonal communication, and collective decision-making, which are only partially represented in written artefacts and therefore less directly supported by text-based feedback systems. This difference may explain why the intervention demonstrated stronger associations with WC and CT than with COL. The feedback system may have functioned as a scaffold by providing timely guidance during task performance, thereby reducing cognitive barriers and facilitating learners' progression toward higher levels of performance. This interpretation is consistent with sociocultural learning theory and the ZPD and is further supported by recent studies investigating AI-assisted scaffolding in higher education learning environments [31].

The results are also consistent with previous studies on technology-enhanced formative assessment reporting positive associations between AI-supported formative feedback systems and improvements in writing quality, argumentation, and revision behavior [32]. Prior research on automated writing evaluation systems similarly suggests that structured feedback may support improvements in coherence, organization, and evidence integration during revision cycles [33]. The substantial improvements observed in WC may additionally reflect the close alignment between rubric-based feedback and structured writing tasks implemented throughout the intervention. Research on AI-assisted writing support has likewise reported that this type of formative feedback tends to be more effective when learners engage in repeated revision processes guided by explicit performance criteria [34]. Similarly, improvements in CT align with prior evidence suggesting that structured feedback can strengthen analytical reasoning and evidence-based argumentation within STEM learning environments [35].

Improvements in COL, although positive, were smaller compared to WC and CT. The observed differences may be interpreted from a sociocultural learning perspective [36], which conceptualizes competency development as a socially mediated process. Previous research in STEM education similarly suggests that collaborative competencies often require more extensive interactional experiences that extend beyond text-based formative assessment mechanisms [37]. This finding further supports earlier observations that effective COL requires richer social interaction and sustained peer engagement than can be provided through automated text-based feedback alone [38].

From a practical perspective, this approach may function as a complementary instructional support mechanism that provides immediate and structured guidance while maintaining instructor oversight and pedagogical support. This interpretation is consistent with recent literature emphasizing that AI should complement rather than replace teacher expertise, particularly in formative assessment contexts where instructional judgment remains essential [39]. The proposed framework may also facilitate the scalable implementation of competency-based formative assessment in large-scale instructional settings. Because the intervention relies on competency descriptors, structured feedback rules, and NLP-assisted analysis rather than discipline-specific content alone, it may be adapted to other interdisciplinary STEM contexts, including mathematics, computer science, applied sciences, and technology-oriented courses that require communication, argumentation, and problem-solving competencies [40].

4.1. Limitations

This study has several limitations that should be considered when interpreting the findings. The quasi-experimental nonequivalent groups design limits causal inference, as intact classroom groups were used and random assignment was not feasible. Although baseline equivalence was established, potential unobserved contextual and individual differences may still have influenced the observed outcomes. The study was conducted within a single institutional and disciplinary context (engineering students at one university in Kazakhstan), which may limit the generalizability of the findings to other educational settings, disciplines, or cultural contexts. The study focused on short-term competency development over an eight-week intervention period, which does not allow conclusions regarding the durability or transferability of the observed effects over time. Future longitudinal research is required to examine the persistence of competency development beyond the intervention period.

4.2. Practical implications

Based on the findings, STEM instructors may consider the following implementation sequence: i) define competency-based assessment criteria using transparent rubrics; ii) design communication-focused STEM tasks aligned with learning outcomes; iii) integrate NLP-assisted formative feedback during iterative learning activities; iv) require revision and resubmission based on feedback; v) monitor progress through instructor validation and feedback analytics; and vi) conduct final competency evaluation using human assessors.

5. CONCLUSION

This study contributes to the growing body of research on AI-supported learning by demonstrating the potential value of integrating structured formative feedback within competency-based STEM education. The findings support the use of AI-assisted formative assessment as a complementary instructional mechanism that can facilitate continuous competency development while preserving the central role of instructor guidance and pedagogical mediation. The study also has practical implications for higher education institutions undergoing digital transformation and competency-based curriculum reform.

AI-supported formative feedback systems may assist educators by providing timely, structured, and scalable support for competency-oriented learning processes, particularly in writing-intensive STEM contexts.

Future research should examine the long-term implementation of AI-supported formative assessment across diverse educational contexts and larger populations. Further studies may also investigate additional competency domains and explore how AI-assisted feedback systems can be integrated with collaborative and socially mediated learning processes to support broader competency development in STEM education.

FUNDING INFORMATION

This study received no specific financial support.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Dilyara Dulatkyzy	✓		✓		✓			✓		✓		✓	✓	
Gulden Akbayeva		✓		✓		✓		✓		✓	✓			✓
Kamalbek Berkimbayev	✓		✓		✓		✓			✓	✓		✓	
Murat Mirzakhmetov	✓		✓			✓		✓	✓		✓		✓	
Nuraim Ibragimova		✓		✓	✓		✓			✓		✓		✓
Saniya Nurgaliyeva	✓		✓			✓		✓	✓		✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The corresponding author may provide study data upon reasonable request.

REFERENCES

[1] T. S. Chu and M. Ashraf, "Artificial intelligence in curriculum design: a data-driven approach to higher education innovation," *Knowledge*, vol. 5, no. 3, p. 14, Jul. 2025, doi: 10.3390/knowledge5030014.
[2] W. Strielkowski, V. Grebennikova, A. Lisovskiy, G. Rakhimova, and T. Vasileva, "AI-driven adaptive learning for sustainable educational transformation," *Sustainable Development*, vol. 33, no. 2, pp. 1921–1947, Apr. 2025, doi: 10.1002/sd.3221.
[3] A. M. H. Chen, E. L. Kleppinger, M. D. Churchwell, and D. H. Rhoney, "Examining competency-based education through the lens of implementation science: a scoping review," *American Journal of Pharmaceutical Education*, vol. 88, no. 2, p. 100633, Feb. 2024, doi: 10.1016/j.ajpe.2023.100633.
[4] P. Black and D. Wiliam, "Developing the theory of formative assessment," *Educational Assessment, Evaluation and Accountability*, vol. 21, no. 1, pp. 5–31, Feb. 2009, doi: 10.1007/s11092-008-9068-5.
[5] B. Thornhill-Miller *et al.*, "Creativity, critical thinking, communication, and collaboration: assessment, certification, and promotion of 21st century skills for the future of work and education," *Journal of Intelligence*, vol. 11, no. 3, p. 54, Mar. 2023, doi: 10.3390/jintelligence11030054.
[6] W. Sun and B. Zhong, "Integrating reading and writing with STEAM/STEM: a systematic review on STREAM education," *Journal of Engineering Education*, vol. 113, no. 4, pp. 939–958, Oct. 2024, doi: 10.1002/jee.20569.
[7] K. A. Callwood, M. Weiss, R. Hendricks, and T. G. Taylor, "Acknowledging and supplanting white supremacy culture in science communication and STEM: the role of science communication trainers," *Frontiers in Communication*, vol. 7, p. 787750, Feb. 2022, doi: 10.3389/fcomm.2022.787750.
[8] G. Rizakhojayeva, S. Ramankulov, M. Akeshova, M. Nurizinova, Y. Tuyakov, and R. Abdrakhmanov, "STEM-based approaches to soft skills development: a synthesis of meta-analytic findings and empirical evidence," *Frontiers in Education*, vol. 10, p. 1663155, Sep. 2025, doi: 10.3389/feduc.2025.1663155.
[9] S. Aziz, "Soft skills development methods in higher education: a systematic literature review," *Higher Education, Skills and Work-based Learning*, vol. 15, no. 4, pp. 868–883, May 2025, doi: 10.1108/HESWBL-02-2025-0050.
[10] R. M. Potluri, M. Gabdullin, A. K. Jumasseitova, and Y. B. Mukashev, "Artificial intelligence–enabled teaching: insights from Kazakhstan higher education students," *Social Sciences and Humanities Open*, vol. 13, p. 102660, Jun. 2026, doi: 10.1016/j.ssaho.2026.102660.
[11] F. Ouyang, T. A. Dinh, and W. Xu, "A systematic review of AI-driven educational assessment in STEM education," *Journal for STEM Education Research*, vol. 6, no. 3, pp. 408–426, Dec. 2023, doi: 10.1007/s41979-023-00112-x.

[12] M. N. Alam, M. A. Islam, M. O. A. Babiker, M. S. Siddiqui, M. bin Amin, and J. Oláh, "AI-assisted learning tools and student learning outcomes: a cognitive load theory perspective," *Computers in Human Behavior Reports*, vol. 21, p. 100986, Mar. 2026, doi: 10.1016/j.chbr.2026.100986.

[13] A. H. Dang, V. S. Huynh, D. D. Nguyen, and Y. Baek, "Beyond automation: cultivating self-regulated learning through AI and reflective practice," *Technology, Knowledge and Learning*, pp. 1–26, Mar. 2026, doi: 10.1007/s10758-026-09970-x.

[14] M. Pikhart and L. H. Al-Obaydi, "Reporting the potential risk of using AI in higher education: subjective perspectives of educators," *Computers in Human Behavior Reports*, vol. 18, p. 100693, May 2025, doi: 10.1016/j.chbr.2025.100693.

[15] S. Wang, F. Wang, Z. Zhu, J. Wang, T. Tran, and Z. Du, "Artificial intelligence in education: a systematic literature review," *Expert Systems with Applications*, vol. 252, p. 124167, Oct. 2024, doi: 10.1016/j.eswa.2024.124167.

[16] M. Abrar, M. ur Rehman, S. Khalid, and R. Ullah, "The intersection of artificial intelligence and assistive technologies in the diagnosis and intervention of mental health conditions," *Artificial Intelligence Review*, vol. 59, no. 2, p. 40, Jan. 2026, doi: 10.1007/s10462-025-11447-9.

[17] D. Yao and J. Lin, "Cognitive enhancement through competency-based programming education: a 12-year longitudinal study," *Education and Information Technologies*, vol. 30, no. 14, pp. 20347–20383, Sep. 2025, doi: 10.1007/s10639-025-13582-w.

[18] H. Y. Durak and A. Onan, "A systematic review of AI-based feedback in educational settings," *Journal of Computational Social Science*, vol. 8, no. 4, p. 96, Nov. 2025, doi: 10.1007/s42001-025-00428-1.

[19] M. Alfaleh, "Sustainable AI-driven assessment in higher education: a systematic review of fairness, transparency, pedagogical innovation, and governance," *Sustainability*, vol. 18, no. 2, p. 785, Jan. 2026, doi: 10.3390/su18020785.

[20] W. N. Alwakid, N. A. Dahri, M. Humayun, and G. N. Alwakid, "Exploring the role of AI and teacher competencies on instructional planning and student performance in an outcome-based education system," *Systems*, vol. 13, no. 7, p. 517, Jun. 2025, doi: 10.3390/systems13070517.

[21] L. H. Alghamdi and T. M. Alghizzi, "Integrating synergistic learning and ability grouping with AI-generated multi-tiered tasks to enhance reading attitudes and outcomes," *Acta Psychologica*, vol. 263, p. 106188, Mar. 2026, doi: 10.1016/j.actpsy.2025.106188.

[22] S. Kohnke and T. Zaugg, "Artificial intelligence: an untapped opportunity for equity and access in STEM education," *Education Sciences*, vol. 15, no. 1, p. 68, Jan. 2025, doi: 10.3390/educsci15010068.

[23] A. Mukashova *et al.*, "Correction: AI-driven framework for automated competency formalization: from professional standards to adaptive learning outcomes," *Frontiers in Computer Science*, vol. 8, p. 1798362, Feb. 2026, doi: 10.3389/fcomp.2026.1798362.

[24] M. Kashif, M. Ammar, A. Sellami, T. K. F. Chiu, S. A. Abbasi, and Z. Ahmad, "Teachers' perspectives on AI integration in K-12 education: challenges, opportunities, and preliminary assessment model–a systematic review," *Computers in the Schools*, pp. 1–27, Dec. 2025, doi: 10.1080/07380569.2025.2602507.

[25] A. Nazyrova, M. Miłosz, G. Bekmanova, A. Omarbekova, G. Aimicheva, and Y. Kadyr, "The digital transformation of higher education in the context of an AI-driven future," *Sustainability*, vol. 17, no. 22, p. 9927, Nov. 2025, doi: 10.3390/su17229927.

[26] Z. Zhumabayeva, K. Zhaxylikova, A. Omirzakova, E. Aitenova, M. Zhailauova, and S. Nurgaliyeva, "Enhancing teaching skills through digital feedback in microteaching: a study with prospective primary teachers," *International Journal of Information and Education Technology*, vol. 15, no. 9, pp. 1820–1828, 2025, doi: 10.18178/ijiet.2025.15.9.2383.

[27] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, "Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses," *Behavior Research Methods*, vol. 41, no. 4, pp. 1149–1160, Nov. 2009, doi: 10.3758/BRM.41.4.1149.

[28] Association of American Colleges and Universities (AAC&U), "VALUE: authentic evidence of student learning." Accessed: Feb. 10, 2026. [Online]. Available: https://www.aacu.org/value

[29] W. Tao, "Fostering inspirational learning through AI-enhanced formative assessment: strategies and challenges in higher education," in *Educational Assessments in the Age of Generative AI*, P. W. Wachira, X. Liu, and S. Koc, Eds., Hershey, PA: IGI Global Scientific Publishing, 2024, pp. 207–249, doi: 10.4018/979-8-3693-6351-5.ch008.

[30] F. Naseer, M. U. Khalid, N. Ayub, A. Rasool, T. Abbas, and M. W. Afzal, "Automated assessment and feedback in higher education using generative AI," in *Transforming Education with Generative AI: Prompt Engineering and Synthetic Content Creation*, R. C. Sharma and A. Bozkurt, Eds., Hershey, PA: IGI Global Scientific Publishing, 2024, pp. 433–461, doi: 10.4018/979-8-3693-1351-0.ch021.

[31] L. Cai, M. M. Msafiri, and D. Kangwa, "Exploring the impact of integrating AI tools in higher education using the zone of proximal development," *Education and Information Technologies*, vol. 30, no. 6, 2025, doi: 10.1007/s10639-024-13112-0.

[32] I.-G. Kang, R. Moussavi, C. E. Castle, C. Pittack, and S. R. McEathron, "A scoping umbrella review of competency-based education: part II – toward a conceptual meta-framework through a synthesis of theories, models, and frameworks," *Competency-Based Education Research Journal*, vol. 3, no. 1, pp. 1–42, Feb. 2026, doi: 10.17161/cberj.v3no1.24119.

[33] I. A. Panhwar, M. Usman, F. Panhwar, and T. A. Surahio, "Exploring the impact of zone of proximal development and scaffolding in second language acquisition: a comparative study of Vygotskian and Freirean approaches," *Social Science Review Archives*, vol. 3, no. 1, pp. 1166–1176, Feb. 2025, doi: 10.70670/sra.v3i1.416.

[34] G. Trajkovski and H. Hayes, "AI-assisted formative assessment and feedback," in *AI-Assisted Assessment in Education: Transforming Assessment and Measuring Learning*, G. Trajkovski and H. Hayes, Eds., Cham: Palgrave Macmillan, 2025, pp. 283–312, doi: 10.1007/978-3-031-88252-4_7.

[35] T. N. Hopfenbeck, Z. Zhang, S. Z. Sun, P. Robertson, and J. A. McGrane, "Challenges and opportunities for classroom-based formative assessment and AI: a perspective article," *Frontiers in Education*, vol. 8, p. 1270700, Nov. 2023, doi: 10.3389/feduc.2023.1270700.

[36] D. Tzuriel, "The socio-cultural theory of Vygotsky," in *Mediated Learning and Cognitive Modifiability. Social Interaction in Learning and Development*, D. Tzuriel, Ed., Cham: Springer International Publishing, 2021, pp. 53–66, doi: 10.1007/978-3-030-75692-5_3.

[37] G. Helm and F. Hesse, "Revising with AI: aligning feedback technologies with learners' revision processes," *Frontiers in Education*, vol. 11, p. 1702406, Feb. 2026, doi: 10.3389/feduc.2026.1702406.

[38] F. Hasançebi, Ö. Güner, C. Kutru, and M. Hasançebi, "Impact of STEM integrated argumentation-based inquiry applications on students' academic success, reflective thinking and creative thinking skills," *Participatory Educational Research*, vol. 8, no. 4, pp. 274–296, Dec. 2021, doi: 10.17275/per.21.90.8.4.

[39] C. Y. Ersanlı and D. B. A. Yeşilel, "The use of automated writing evaluation tools to foster ESL writing instruction," in *New Directions in Technology for Writing Instruction. English Language Education*, G. Yangın-Ekşi, S. Akayoglu, and L. Anyango, Eds., Cham: Springer International Publishing, 2022, pp. 193–209, doi: 10.1007/978-3-031-13540-8_10.

[40] H. Ren, W. Zhang, and H. Cao, "Formative assessment as a self-regulation engine in blended language learning: a case study with top-tier engineering talents," *Acta Psychologica*, vol. 264, p. 106505, Apr. 2026, doi: 10.1016/j.actpsy.2026.106505.

BIOGRAPHIES OF AUTHORS



Dilyara Dulatkyzy is a 2nd-year Ph.D. student in the "Foreign Language: Two Foreign Languages" educational program at Karaganda National Research University named after Academician Ye. A. Buketov. She holds a master's degree in the same field and has a strong background in digital transformation. Her research interests include information technologies in higher education, multilingual education, and the digitalization of foreign language teaching. She has contributed as a junior researcher to state-funded scientific projects, specifically focusing on pedagogical training for inclusive environments. She can be contacted at email: dileeka_b@bk.ru.



Gulden Akbayeva is a candidate of Pedagogical Sciences and an associate professor at the Department of Theory and Methods of Foreign Language Training at Karaganda National Research University named after academician Ye. A. Buketov. Her professional experience includes serving as the head of the Department of English Language and Linguodidactics (2013–2018). Her research interests focus on innovations in linguodidactics, digitalization of foreign language education, content and language integrated learning (CLIL), and the implementation of the Bologna process parameters. She can be contacted at email: rgul.ksu@mail.ru.



Kamalbek Berkimbayev is a Doctor of Pedagogical Sciences and a Professor in the Department of Pedagogical Sciences at Khoja Akhmet Yassawi International Kazakh-Turkish University. His research interests encompass professional education, innovations in teaching, and the informatization of education. He can be contacted at email: kamalbek.berkimbaev@ayu.edu.kz.



Murat Mirzakhmetov received his candidate of Pedagogical Sciences degree from the Khoja Akhmet Yassawi International Kazakh-Turkish University (Kazakhstan). He has over 50 years of experience working at the Abai Kazakh National Pedagogical University, where he is currently a senior lecturer at the Department of Russian Language and Literature in the Faculty of Philology. His current research interests include student learning and development at various levels and in different fields of education. His publications cover topics such as teaching methodology and educational courses. He can be contacted at email: m.mirzakhmetov@abaiuniversity.edu.kz.



Nuraim Ibragimova received her degree in Mathematics and Physics from the Gogol Pedagogical Institute in 1988 and completed postgraduate studies at the Novosibirsk State University (1993–1996). She obtained her Ph.D. in Education in 2007. She has extensive experience in higher education and previously served as Vice-Rector for Academic Affairs and Research at Bolashak University. She is currently a senior lecturer in Informatics and ICT. Her research interests include computer science education, assessment of learning outcomes, and educational methodologies. She can be contacted at email: Nurym1966@gmail.com.



Saniya Nurgaliyeva is a candidate of Pedagogical Sciences and a Distinguished University Professor at Abai Kazakh National Pedagogical University. She is currently participating in a Visiting Professor Program at Ataturk University in Turkey. The winner of the republican competition of the Ministry of Education and Science among higher education institutions of the Republic of Kazakhstan "The best teacher of higher education institution" – 2017. Her research focuses on innovative processes in education and national upbringing, professional development models and their impact on student learning, as well as the implementation and effects of software environments in the classroom. She can be contacted at email: sanianurgalieva77@gmail.com.