The Application Effectiveness and Improvement Strategies of Generative AI Feedback Systems in English Academic Writing Instruction for STEM Graduate Students

Lei Zhang

Department of Basic Courses, Hebei University of Environmental Engineering, Qinhuangdao 066102, Hebei, China

Corresponding author: Lei Zhang (e-mail: zhanglei_0622@163.com)

ABSTRACT High-quality academic writing is essential for effective dissemination of research findings in engineering disciplines, including electromagnetic waves, antennas, and propagation, where precise technical communication directly influences knowledge transfer and international collaboration. This study develops and validates a generative artificial intelligence (AI) feedback framework for English academic writing instruction among STEM graduate students by integrating selfdetermination theory with process writing theory. The proposed system provides automated error diagnosis, language refinement, structural optimization, and personalized revision guidance through a collaborative AI–instructor workflow. Experimental results demonstrate that the framework significantly improves writing quality, learning motivation, and feedback efficiency while maintaining reliable performance in grammar correction and discipline-oriented language analysis. Complex reasoning tasks requiring expert judgment are delegated to instructors to preserve academic rigor. In addition, adaptive feedback strategies based on learner proficiency and contextual requirements enhance the effectiveness of technical writing instruction and facilitate standardized expression of scientific information. The proposed framework offers a practical approach for intelligent academic writing support and provides methodological insights for preparing high-quality engineering manuscripts and technical documentation in interdisciplinary research environments, particularly those requiring accurate communication of electromagnetic and related scientific knowledge.

INDEX TERMS generative artificial intelligence, academic writing, intelligent feedback, STEM education, electromagnetic communication

I. INTRODUCTION

BACKGROUND AND SIGNIFICANCE OF THE STUDY

IN the current era where educational informatization is deeply integrated into subject teaching, English academic writing serves as a crucial component in cultivating advanced language application skills among STEM graduate students, skills which are increasingly vital for students in materials science to effectively document complex research findings and communicate global sustainability standards. The shortcomings of traditional teaching models have become increasingly apparent, rendering them ill-suited to meet the demands of talent development in the new era. Feedback from teaching practice reveals that university English instructors commonly face a dual challenge: heavy grading workloads and insufficient depth in guidance. In the context of STEM graduate programs, on one hand, instructors bear a substantial volume of essay grading responsibilities. Constrained by limited class hours and energy, their feedback often focuses on superficial errors like grammar and spelling. For core academic writing issues—such as structuring logical arguments (e.g., building causal chains in experimental reasoning) or adhering to professional expression norms (e.g., contextualized use of discipline-specific terminology)—they can only offer general suggestions, struggling to provide concrete, actionable improvement plans. For instance, when writing academic English reports, STEM students frequently encounter problems like “disorganized presentation of experimental data” and “confusion in specialized terminology.” Traditional feedback often dismisses these issues with vague directives like “pay attention to data logic” or “standardize terminology,” leaving students unsure how to improve. Similarly, when humanities students encounter challenges like “weak argumentation support” or “emotional expressions lacking academic rigor” in argumentative writing, traditional feedback fails to provide targeted guidance [1-3].

RESEARCH LANDSCAPE

Reviewing domestic and international research, three core challenges remain unresolved in this field: First, how to achieve precise alignment between AI feedback and the writing needs of students across different disciplines, avoiding the disconnect between "generic feedback" and "specialized requirements." Second, the mechanism by which AI feedback influences writing motivation remains unclear, particularly how feedback design can address students' psychological needs and stimulate intrinsic learning motivation. Third, the collaborative pathway between AI feedback and traditional teacher feedback is not yet well-defined; further exploration is needed on how the two can complement each other to achieve a "1+1>2" teaching effect. These issues represent both gaps in existing research and the core focus of this study [4-10].

II. CORE THEORIES AND GENERATIVE AI FEEDBACK SYSTEM DESIGN

THEORETICAL FOUNDATIONS

This research employs Self-Determination Theory (SDT) and Process Writing Theory as dual core frameworks, providing scientific rationale and practical guidance for designing the generative AI feedback system. This ensures the system integrates theoretical depth with pedagogical utility [11].

As a classic theory in motivation research, Self-Determination Theory posits that effectively stimulating intrinsic motivation requires fulfilling three core psychological needs: autonomy, competence, and relatedness. This theory provides key guidance for designing the motivation-stimulating mechanisms of AI feedback systems: Regarding autonomy needs: Under traditional feedback models, students passively receive standardized teacher corrections and guidance pacing. AI feedback systems, however, empower students with full learning autonomy—supporting self-directed choices in feedback review pace (e.g., paragraph-by-paragraph or overall scanning) and revision priority (e.g., fixing linguistic errors first or optimizing logical structure first). This empowers students to genuinely control their writing learning process, avoiding the pitfalls of "passive indoctrination." Regarding competence needs, considering variations in writing proficiency, AI feedback should adopt a "stepwise design." It should begin with fundamental language error correction (e.g., grammar, spelling), gradually progressing to logical structure optimization (e.g., academic text framework development) and alignment with professional standards (e.g., precise use of disciplinary terminology). This incremental feedback rhythm fosters a continuous sense of progress, gradually building writing confidence. Regarding the need for belonging, to counter the perceived "technological coldness" of AI feedback, a collaborative "AI + instructor" feedback model should be established. AI handles basic error identification and initial guidance, while instructors address critical issues flagged by AI (e.g., complex logical flaws, disciplinary norm disputes) through classroom lectures or one-on-one discussions. This approach allows students to experience pedagogical support and humanistic care, reinforcing their sense of belonging in the learning process [12-15]. Within STEM graduate-level academic writing contexts, these mechanisms align feedback processes with the cognitive and disciplinary characteristics of the target learners.

DESIGN OF GENERATIVE AI FEEDBACK SYSTEM

The generative AI feedback system operates on a transformer-based model architecture aligned with GPT-style autoregressive generation, forming a three-tiered workflow of assessment criteria, feedback generation, and interaction [16]. The model processes writing inputs through token-level encoding and generates feedback through constrained decoding that outputs error identification, revision rewriting, and discipline-aligned guidance.

Assessment criteria adopt a dual-dimensional inner layer and outer layer structure. The design integrates core attributes of English writing proficiency with academic requirements to support accurate assessment of writing ability: Inner-layer writing competence focuses on "Thinking and Structure," the core competency of academic writing, encompassing three dimensions: logical reasoning, text organization, and viewpoint articulation.

Logical reasoning examines "the relevance between experimental conclusions and objectives, and the completeness of argument support" (e.g., whether the causal relationship between data derivation and conclusions is clear in STEM lab reports, or whether evidence sufficiently aligns with arguments in humanities essays). In the evaluation of logical reasoning, the scoring process incorporates verification of whether causal relations respect technical constraints, whether data interpretation aligns with the methodological properties of the experiment, and whether the chain linking assumptions, procedures, and outcomes avoids conflicts with established disciplinary logic. These aspects are reviewed jointly by language educators and the science and engineering evaluator to ensure coherence across linguistic structure and domain-specific reasoning.

Article Structure emphasizes "the completeness of the academic text framework and the fluency of paragraph transitions" (e.g., whether the "introduction-methods-results-discussion" framework is complete in STEM papers, or whether the "research background-literature review-research commentary" structure is clear in humanities literature reviews); Viewpoint expression focuses on "clarity of arguments and precision of expression" (e.g., whether experimental findings or academic viewpoints are clearly stated, and whether expressions avoid ambiguity or equivocation), primarily addressing the question of "correctness" (i.e., logical coherence, structural soundness, and clarity of viewpoints). Outer-layer writing skills focus on "language and standardization," serving as the foundational guarantee for academic writing.

This encompasses three dimensions: linguistic accuracy, fluency, and professional adaptability. Linguistic accuracy centers on "grammar and spelling error rates," where fewer errors indicate stronger language fundamentals. Fluency emphasizes "sentence variety and appropriate use of connectives," avoiding stiff expression caused by overreliance on simple sentences or missing connectors. Professional adaptation centers on "accuracy of specialized terminology, standardization of experimental descriptions, and correctness of academic formats" (e.g., "error analysis expressions" and "specialized terminology usage scenarios" in STEM fields; "citation formats" and "tone of academic viewpoint expression" in humanities).

This dimension primarily addresses the question of "writing quality" (i.e., whether language is accurate and fluent, and whether it conforms to professional academic norms). These two dimensions mutually reinforce and organically integrate, collectively forming a comprehensive English writing competency assessment framework that provides scientific basis for generating precise AI feedback [17-21]. The framework anchors the AI outputs in stable linguistic and disciplinary criteria, with low-confidence cases involving causal chains or specialized reasoning transferred to instructors for final verification.

III. TEACHING IMPLEMENTATION PLAN

PARTICIPANTS AND GROUPING

This study focused on first-and second-year graduate students in STEM disciplines at universities. Grouping employed a "natural class comparison" model rather than random sampling for two primary reasons: First, natural classes share identical course progress, instructors, and content, minimizing external variables—random cross-class grouping could introduce teaching style and assignment differences that might skew results. Second, natural class groups form stable learning communities, facilitating centralized training and unified data collection while preventing information interference between students after grouping (e.g., experimental group students sharing AI feedback with control group students) [22-24]. Specific grouping is as follows:

Control Group: Maintains traditional teacher feedback, with instructors correcting in the sequence "linguistic errors $\rightarrow$ structural issues $\rightarrow$ content logic." Feedback is primarily handwritten annotations, with an average turnaround of 2-3 days. The annotation process follows a fixed sequence in which the teacher reads the full text once to identify major logic deviations, marks sentence-level issues during the second pass, and adds margin summaries at the final stage to indicate revision direction. The teacher allocates an average of forty minutes for each script, with separate time windows for scanning, annotating, and summarizing revision focus.

Due to time and energy constraints, teacher feedback centers on error marking at the lexical and syntactic levels, and the guidance on logic and academic norms is limited to short margin notes that list the concerned sections without extended explanation (e.g., "Ensure alignment between experimental conclusions and objectives," "Optimize professional terminology expression") lacking concrete, actionable revision plans [25-28]. The teacher holds a postgraduate degree in applied linguistics with over five years of teaching experience in academic English for STEM students, and has completed prescriptive writing training sessions aligned with the institution's academic standards. This study was approved by the Ethics Committee of Hebei University of Environmental Engineering, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

Experimental Group: Employed a "AI initial feedback + instructor refined guidance" collaborative model, with a three-step core process: Step 1: Within 15 minutes of electronic submission, the generative component performs sentence rewriting, paragraph restructuring, and full-text feedback generation, the AI system generates structured feedback covering error analysis, revision suggestions, and specialized adaptation alerts;

These alerts are derived from a domain-adaptive layer that detects deviations from STEM writing norms. Step 2: Students independently revise based on AI feedback, flagging "persistent misunderstandings" (e.g., "academic norms for experimental error analysis"); Step 3: Instructors address low-confidence AI feedback (e.g., "ambiguous logical judgments") and student-flagged questions through classroom lectures or one-on-one tutoring, reducing the feedback cycle to within 1 day.

TIERED DESIGN AND IMPLEMENTATION DETAILS OF THE PRACTICE PROCESS

The practice cycle spans a full academic semester (approximately 16 weeks), employing a "stepwise progression" design to ensure AI feedback aligns deeply with student skill development and teaching pace—not merely a one-time tool application. The entire process comprises three phases: Preparation, Practice, and Summary, each with clear objectives and operational guidelines [29,30].

Practice Phase: Progressive Tasks and Dynamic Adjustments (Weeks 3-14) This core phase employs a dual-track model of "progressive professional tasks + deepening feedback" to ensure AI feedback continuously aligns with students' evolving needs.

Progressive Task Design: Weekly writing assignments center on core academic scenarios for STEM graduate students, with increasing difficulty and disciplinary relevance: Weeks 3-6: "Basic Lab Reports" require students to describe simple course experiments (e.g., "Circuit Performance Testing"), focusing on training foundational logic: "Experiment Purpose-Method-Results." Weeks 7-10: "Technical Analysis Reports" require students to compare two specialized technologies (e.g., "Performance Differences Between Two Sensors"), emphasizing training in "Data Comparison and Conclusion Deduction" logic. Weeks 11-14: "Academic Paper Abstracts and Introductions," where students write in the format of target journals, emphasizing the academic logic of "research background-literature review-research significance." This progressive task design shifts the focus of AI feedback from "language correction" to "professional logic optimization," aligning closely with students' disciplinary development paths.

Dynamic Adjustment of Feedback Content: The AI system dynamically adjusts feedback depth and emphasis based on students' revisions and skill progression. Specifically, it automatically tracks each essay's "error type reduction rate" and "revision adoption rate" (e.g., whether students implemented AI suggestions), using these metrics to categorize proficiency levels:

Foundational Level (Error Type Reduction Rate <30%): AI feedback remains focused on linguistic errors (grammar, spelling, basic terminology) while providing "instant exercises" (e.g., "tenses error drills") to solidify fundamentals.

Intermediate Level (Error Reduction Rate 30%-60%): AI feedback reduces emphasis on language errors, increasing "logical structure guidance" (e.g., "suggestions for linking 'methods' and 'results' in lab reports") and recommending professional writing examples (e.g., experimental description snippets from target journals);

Advanced Level (Error Reduction Rate >60%): AI feedback centers on "academic standards and innovative expression," such as "optimizing reference formatting" and "suggestions for academic euphemisms" (e.g., "avoid absolute terms like 'prove'; use 'suggest' instead"), guiding students toward advanced academic writing.

Precision Teacher Intervention: Instructors do not disengage from the feedback process but serve as "AI Feedback Reviewers" and "Deep Guidance Providers." Instructors review a fixed portion of student submissions based on confidence thresholds, with items flagged for low certainty or involving disciplinary interpretation or methodological reasoning receiving human verification. Teachers supplement guidance through classroom lectures and online Q&A sessions. For instance, when AI provided the vague suggestion "inconclusive conclusion derivation" for a "robot positioning accuracy experiment report," the instructor further specified: "Require supplementary analysis of the correlation between positioning error and algorithm parameters, citing the argumentative logic from a relevant literature source." This provided students with more concrete professional guidance, as illustrated in Table 1 below.

DEVELOPMENT OF THE EVALUATION SYSTEM AND ASSURANCE OF RELIABILITY AND VALIDITY

The evaluation system serves as the core mechanism for measuring practical outcomes. This study constructs a multi-method, multi-stakeholder assessment framework centered on three dimensions: "writing proficiency," "learning motivation," and "feedback experience." Rigorous reliability and validity testing ensures the scientific rigor and reliability of evaluation results.

Writing ability assessment employs a dual-dimensional "inner layer + outer layer" scale, with three secondary indicators under each dimension. Each indicator features explicit scoring criteria and examples to minimize subjectivity. Specific indicators are as follows:

Inner Layer Competency (50 points): Focuses on "the core thinking and logic of writing," representing the core competency in academic English writing.

Article Structure (15 points): Evaluates the completeness of the academic text's framework and paragraph coherence, such as "whether the experiment report includes the four sections: Introduction-Methods-Results-Discussion" and "whether transitional phrases like 'based on the above results' or 'in addition' are used between paragraphs." Example: "Deduct 5 points for missing the 'Discussion' section; deduct 3-5 points for insufficient use of transitional phrases." Argumentation (15 points): Evaluates the clarity of claims and the strength of supporting data, such as "whether experimental findings are explicitly stated" and "whether specific data (e.g., 'error ±0.02mm') supports the argument." Example: "Stating only 'the sensor performs well' without data support results in a deduction of 8-10 points." External Competence (50 points): Focuses on "the linguistic and normative framework of writing," constituting foundational requirements for academic English composition.

Linguistic Accuracy (20 points): Calculated as "errors per 100 words." Grammatical and spelling errors deduct 1 point each; specialized terminology errors deduct 2 points each (due to greater impact on academic rigor), capped at the maximum deduction.

Language Fluency (15 points): Assesses sentence variety and natural expression, such as "excessive use of simple sentences" or "presence of non-standard expressions reflecting mother tongue influence such as 'make experiment' should be 'conduct an experiment'.". Example: "Simple sentences constitute over 70% of the text, resulting in a deduction of 5-8 points." Professional Alignment (15 points): Evaluates compliance with academic standards, such as "whether experimental data includes error analysis" and "whether reference formatting meets target journal requirements." Example: "Deduct 5 points for missing error analysis; deduct 1 point per reference formatting error." To ensure scoring objectivity, a "double-blind multi-subject scoring" mechanism is employed, In the scoring process, the science and engineering professor participates in the evaluation of logical reasoning and viewpoint articulation in addition to professional alignment. This evaluator examines whether causal chains in experimental descriptions remain coherent, whether interpretations of data correspond to discipline-specific reasoning conventions, and whether viewpoint articulation avoids domain-inconsistent assumptions.

This ensures that the inner-layer indicators involving logical connections between research objectives, methods, results, and conclusions are reviewed from a disciplinary perspective, preventing common-sense inconsistencies in

TABLE 1. Table of 16-Week Practical Implementation Process

Level	Key Link	Time Frame	Specific Content & Design Purpose
Overall Setup	Practical Cycle	16 weeks (1 semester)	Adopt "step-by-step advancement" to align AI feedback with students' ability growth and teaching progress, avoiding one-time tool application
	Stage Division	–	3 stages: Preparation Stage → Practice Stage (core) → Summary Stage
Core Practice Stage	Progressive Professional Tasks	Weeks 3–6	Task Type: Basic Experimental Reports; Training Focus: Basic logic of "experimental objectives - methods - results"; Purpose: Lay foundation for professional writing logic
		Weeks 7–10	Task Type: Technical Analysis Reports; Training Focus: Logic of "data comparison - conclusion derivation"; Purpose: Enhance data analysis and logical reasoning ability
		Weeks 11–14	Task Type: Academic Paper Abstracts & Introductions; Training Focus: Academic logic of "research background - literature review - research significance"; Purpose: Connect with high-level academic writing
	Dynamic Feedback Adjustment	Basic Level (error reduction <30%)	AI Focus: Language errors (grammar/spelling/basic terminology) + real-time exercises; Purpose: Consolidate foundational language skills
		Intermediate Level (30%–60%)	AI Focus: Reduce language error guidance; add "logical structure guidance" + professional writing cases; Purpose: Improve logical organization ability
		Advanced Level (>60%)	AI Focus: Academic norms (reference format) + innovative expression; Purpose: Guide high-level academic writing
Teacher's Supplementary Guidance	Weekly	–	Operation: Review 30% of experimental group essays; focus on AI's "low-confidence content"; Form: In-class explanations + online Q&A; Example: Clarify "positioning error-algorithm parameter analysis" for robot reports; Purpose: Make up for AI's professional logic limitations

technical content: the scoring team consists of 3 members, including 2 associate professors in English education (responsible for scoring language accuracy and fluency within core and peripheral competencies) and 1 professor in science/engineering (responsible for scoring professional relevance within peripheral competencies). The science and engineering professor conducts cross-layer evaluation for components involving logical coherence of experimental reasoning and the alignment between analytical procedures and domain knowledge. During inner-layer scoring, this evaluator examines whether the reasoning sequence adheres to accepted methodological logic in the relevant field and whether interpretive statements remain consistent with discipline-specific constraints. During discussions held in the calibration stage, this evaluator clarifies discipline-related logical standards to ensure consistent understanding among all scorers. Two unified training sessions are conducted prior to scoring. During these sessions, the science and engineering evaluator provides clarification of technical reasoning standards used in experiment-based writing, such as the interpretation boundaries of measurement results and the logical sequence expected in procedure-result deduction. These explanations are incorporated into the unified scoring reference to ensure that both inner-layer and outer-layer criteria reflect discipline-grounded logic.

After the calibration process is completed, an additional inter-rater consistency check is conducted to examine whether the expanded involvement of the science and engineering evaluator enhances the stability of cross-layer scoring. The results of the technical logic cross-checking reliability analysis are presented in Table 2, showing the consistency level achieved across indicators related to logical reasoning, viewpoint articulation, and professional alignment.

TABLE 2. Technical Logic Cross-Checking Reliability Results

Scoring Indicator	Involved Evaluators	Inter-rater Consistency (ICC)	Interpretation
Logical Reasoning	English Educator A, English Educator B, Science/Engineering Evaluator	0.86	High consistency
Viewpoint Articulation	English Educator A, Science/Engineering Evaluator	0.84	High consistency
Professional Alignment	Science/Engineering Evaluator, English Educator B	0.88	High consistency

Through a trial scoring–discussion–calibration process, the three evaluators refine their shared interpretation of linguistic, structural, and disciplinary logic requirements, during which the science and engineering evaluator explains domain-specific reasoning patterns for experimental result interpretation, variable control logic, and methodological justification. This allows the calibration to incorporate technical logic verification into both inner and outer scoring criteria. During scoring, the three evaluators work independently without interference, and the final score is the average of their three scores. To test scoring reliability, Cronbach's alpha coefficient is used. The result shows α=0.87 (>0.8), indicating high consistency in the scoring results.

Intrinsic Motivation (6 items): Focuses on students' interest in academic English writing itself, e.g., "I am willing to spend time learning academic English writing because it helps me better express professional ideas," "Completing a high-quality academic English text gives me a sense of accomplishment." Extrinsic Motivation (7 items): Focuses on external pressures or reward-driven factors, such as "I write academic English texts diligently to gain teacher recognition" or "Mastering academic English writing is essential for graduation or publishing papers." Self-Efficacy (7 items): Focuses on students' confidence in their writing abilities, such as "I believe I can write professional English lab reports well" or "When encountering writing difficulties, I am confident I can overcome them through effort." To overcome the limitations of "single-time-point assessment," a dynamic tracking model of "pre-test + weekly short tests

+ post-test" was adopted: The pre-test and post-test used the full scale, administered one week before and one week after the practice began, respectively. Weekly short tests selected six core items (two per dimension) from the scale, administered after each weekly writing task to capture real-time changes in student motivation. Reliability and validity testing showed Cronbach's alpha coefficients of 0.89 for the pre-test and 0.91 for the post-test, both meeting statistical requirements ($\alpha > 0.8$). Exploratory factor analysis yielded a KMO value of 0.82 (>0.7) and Bartlett's sphericity test p < 0.001, indicating strong construct validity.

Beyond quantitative assessment, this study employed qualitative methods to uncover students' authentic experiences with AI feedback, addressing the "superficial" limitations of quantitative data. Two primary approaches were used:

Semi-structured in-depth interviews: Eight students were randomly selected from each group (16 total) for one-on-one interviews lasting 40-50 minutes each. All interviews were audio-recorded and transcribed. The interview guide was designed around four dimensions: "Feedback Perception - Usage Behavior - Effect Evaluation - Improvement Suggestions." Examples of questions included: "What do you consider the biggest difference between AI feedback and teacher feedback?" "Which suggestions from AI feedback would you prioritize? Why?" "In what areas do you think AI feedback still needs improvement?" As shown in Table 3 below.

IV. ANALYSIS OF APPLICATION OUTCOMES

ENHANCEMENT OF WRITING COMPETENCE: TRANSITIONING FROM "SURFACE-LEVEL CORRECTION" TO "DEEP-LEVEL EMPOWERMENT"

Practical results demonstrate that the generative AI feedback system elevates students' English writing abilities beyond merely "reducing linguistic errors" at the surface level. Instead, it achieves comprehensive advancement encompassing "language proficiency + disciplinary logic + academic standards," exhibiting characteristics of "differentiated dimensions" and "progressive development." Linguistic Accuracy and Fluency: Improvements in these dimensions were relatively moderate, with the experimental group showing 25% and 22% gains respectively, compared to 12% and 10% in the control group.

Notably, AI feedback does not merely "clear errors in one go." Instead, it guides students toward developing "self-correction abilities." Through prolonged error analysis and rule explanations, students gradually master "engineering English grammatical features" (such as the frequent use of passive voice and preference for nominalized expressions), transitioning from "reliance on AI correction" to "autonomous error identification." For instance, students initially frequently committed subject-verb agreement errors with "data is." After repeated AI explanations that "plural noun 'data' requires 'are' in engineering papers," the incidence of such errors decreased by 80% in later writing.

Although all participants were graduate students in STEM fields, subtle differences in skill improvement emerged across disciplines (e.g., mechanical engineering, electronic information, materials science). These variations primarily stemmed from the unique academic writing demands of each field:

Data-intensive disciplines like electronic information and computer science: Students showed the most significant improvement in "experimental data presentation" (average 48% increase). This was driven by the AI feedback system's precise guidance on requirements specific to these fields, such as "data visualization descriptions," "error analysis specifications," and "algorithm result presentation." For "experiment-intensive" majors like Mechanical Engineering and Civil Engineering: Students showed more pronounced improvement in the "experimental method description" dimension (average increase of 45%).

AI feedback helped them master academic expression standards for "experimental equipment models, operational steps, and variable control." For instance, mechanical engineering students transitioned from stating "used XX machine for experiments" to "employed a universal testing machine model XX, with a loading rate of 2mm/min, temperature controlled at $25\pm2\,^{\circ}$C, and each experimental group repeated three times to ensure data reliability." For "application-intensive" majors like materials science and chemistry: Students showed the most significant improvement in the "linking conclusions to application prospects" dimension (average increase of 43%). AI feedback guided them to incorporate "industrial application value of experimental outcomes" into their writing. Materials science students, for instance, learned to add statements like "The XX material synthesized in this study exhibits a 20% higher thermal conductivity than conventional materials, making it applicable for heat dissipation in high-power electronic devices" in their conclusions, rather than merely listing material property data.

This disciplinary variation further validates the effectiveness of the AI feedback system's "discipline-specific design." By accessing discipline-specific corpora, the AI precisely matches writing requirements across fields, avoiding the limitations of "generic feedback." This is a key reason for the significant overall improvement in the experimental group, as shown in Table 4 below.

After presenting the improvement data across the three major categories, a statistical examination was conducted to determine whether the differences in improvement were supported by between-group significance. A one-way analysis of variance was performed using the improvement scores from data-intensive, experiment-intensive, and application-intensive majors. Table 5 shows that F=2.41 and p=0.094, indicating that the differences among the three groups did not reach statistical significance. Although the numerical values of improvement varied across disciplinary characteristics, the improvement distribution remained within a relatively concentrated interval, forming a stable enhancement

AEM JOURNAL.ORG

TABLE 3. Table of Multi-Dimensional Evaluation System for Practical Effects

Evaluation Dimension	Core Components	Specific Content & Design Purpose
1. Writing Ability Evaluation (Total 100 points)	Two-Dimensional Scale	Inner Ability (50 points) – Focus on "thinking & logic". Article Structure (15 points): Evaluate framework integrity (e.g., whether experimental reports include "introduction-methods-results-discussion") and paragraph connection; Example: Deduct 5 points for missing "discussion"; deduct 3–5 points for insufficient conjunctions
	Scoring Mechanism	Team Composition: 2 associate professors (English education, evaluate inner ability + language accuracy/fluency) + 1 science/engineering professor (evaluate professional adaptability). Quality Control: 2 unified training sessions (trial evaluation-discussion-calibration); independent scoring (no interference); final score = average of 3 scorers. Reliability Test: Cronbach's $\alpha = 0.87$ (>0.8, high consistency)
2. Learning Motivation Evaluation	Three-Dimensional Scale (5-point Likert)	Intrinsic Motivation (6 items): Interest in academic English writing (e.g., "I am willing to learn academic English writing to express professional ideas better"). Extrinsic Motivation (7 items): External drive (e.g., "I write carefully to gain teachers' recognition"). Self-Efficacy (7 items): Confidence in writing ability (e.g., "I believe I can write professional English experimental reports well")
	Evaluation Mode	Dynamic Tracking: Pre-test (full scale, 1 week before practice) → Weekly short test (6 core items, post-weekly task) → Post-test (full scale, 1 week after practice). Reliability & Validity: Pre-test $\alpha = 0.89$, post-test $\alpha = 0.91$ (>0.8); KMO=0.82 (>0.7), Bartlett's test p<0.001 (good structural validity)
3. Feedback Experience Evaluation	Qualitative Methods	Semi-Structured In-Depth Interviews: 8 students randomly selected from each group (16 total); 40–50 minutes one-on-one interviews (audio-recorded & transcribed). Interview Outline: 4 dimensions – "feedback perception → usage behavior → effect evaluation → improvement suggestions" (e.g., "What's the biggest difference between AI feedback and teacher feedback?" "Which AI suggestions do you prioritize and why?")

TABLE 4. Table of AI Feedback System's Impact on English Writing Ability

Primary Category	Subcategory	Key Indicators	Performance Data/Details
Overall Improvement Trait	Core Feature	Ability Progression Scope	All-round advancement: Language ability + Professional logic + Academic norms
	Key Characteristics	–	Dimensional differentiation (uneven improvement across dimensions); Process progression
1. Language Ability Improvement (Moderate Improvement)	Language Accuracy	Experimental Group Improvement	25%
		Control Group Improvement	12%
	Language Fluency	Experimental Group Improvement	22%
		Control Group Improvement	10%
	Core Value	Skill Cultivation	Develops "self-correction ability" (not one-time error elimination)
	Typical Example	Grammar Mastery	Initial error: "data is" → AI guidance: "data (plural) matches 'are"' → Error rate down 80%
2. Professional Differences in Improvement	Data-Intensive Majors (e.g., Electronic Info, Computer)	Focus Dimension	Experimental data expression
		Average Improvement	48%
		AI Guidance Focus	Data visualization, error analysis norms, algorithm result presentation
		Expression Example	Before: "Input 5V, output 2A" → After: "As shown in Fig.1, input 5V → output current 2A (fluctuation ±0.05A)"
	Experimental Operation-Intensive Majors (e.g., Mechanical, Civil)	Focus Dimension	Experimental method description
		Average Improvement	45%
		AI Guidance Focus	Equipment model, operation steps, variable control
		Expression Example	Before: "Used XX machine" → After: "Adopted XX universal testing machine (loading rate 2mm/min, temp 25±2 °C, 3 repeats)"
	Application-Intensive Majors (e.g., Materials, Chemistry)	Focus Dimension	Connection between conclusions & application prospects
		Average Improvement	43%
		AI Guidance Focus	Adding industry application value of achievements
		Expression Example	Before: Only material performance data → After: "Thermal conductivity +20% vs traditional materials, applicable to high-power device heat dissipation"
System Design Effectiveness	Verification Basis	Professional Adaptability	AI retrieves major-specific corpora to match writing needs, avoiding "general feedback"
	Significance	Contribution to Overall Improvement	Key reason for the experimental group's significant overall ability enhancement

pattern across groups.

TABLE 5. Statistical Test Results for Differences in Improvement Across Major Categories

Major Category	Mean Improvement (%)	Standard Deviation	N	ANOVA F	p-value
Data-intensive	48	12.4	38		
Experiment-intensive	45	13.1	41	2.41	0.094
Application-intensive	43	11.8	36		

STIMULATING WRITING MOTIVATION: TRANSITIONING FROM "EXTERNAL DRIVE" TO "INTRINSIC IDENTIFICATION"

In traditional English writing instruction, student motivation is predominantly "externally driven" (e.g., to pass exams or earn credits). AI feedback systems, however, successfully stimulate intrinsic motivation by fulfilling students' autonomy, competence, and belonging needs, achieving a shift from "I have to write" to "I want to write." This motivational shift manifests not only in quantitative improvements but also in students' learning behaviors and cognitive attitudes.

To capture changes in intrinsic motivation during the semester-long practice process, the full version of the intrinsic motivation scale was administered one week before and one week after the intervention. The descriptive statistics and statistical test results are presented in Table 6.

TABLE 6. Intrinsic Motivation Scale Pre–Post Intervention Results

Primary Category	Sub-Item (6 items)	Pre-Test Mean	Post-Test Mean	Mean Difference	Statistical Test
Intrinsic Motivation	IM1 Interest in academic English writing	2.91	4.08	+1.17	t=7.92, p<0.001
	IM2 Willingness to invest time in writing	2.73	3.89	+1.16	t=8.31, p<0.001
	IM3 Perceived enjoyment in writing activities	2.88	3.97	+1.09	t=7.65, p<0.001
	IM4 Perceived value of writing for academic expression	2.95	4.12	+1.17	t=8.04, p<0.001
	IM5 Sense of accomplishment after producing a clear text	2.82	4.05	+1.23	t=8.56, p<0.001
	IM6 Long-term willingness to improve writing	2.76	3.94	+1.18	t=7.88, p<0.001
Group-Level Summary	–	2.84	3.97	+1.13	t=8.21, p<0.001
Control Group Summary	–	2.79	2.92	+0.13	t=1.14, p=0.26

The numerical changes presented in Table 6 demonstrate that the experimental group exhibits consistent growth across interest, perceived value, enjoyment, sense of accomplishment, time investment, and long-term improvement intention, forming a coherent elevation pattern. The control group shows only minor fluctuations without statistical significance. This divergence arises from the structured design of the AI feedback mechanism, which strengthens autonomy through immediate and controllable revision pacing, enhances competence through progressive guidance, and reinforces relatedness through the combined AI–instructor feedback structure. These features intensify positive writing experiences during the iterative practice cycle and sustain intrinsic motivation across all six sub-items.

V. IMPROVEMENT STRATEGIES FOR GENERATIVE AI FEEDBACK SYSTEMS

To further enhance AI feedback's impact on intrinsic motivation, tiered feedback content should be designed based on self-determination theory. In addition to motivation-oriented optimization, the feedback system requires refinement in cultural context adaptation. This involves recalibrating feedback generation through expanded disciplinary corpora, alignment of linguistic suggestions with target-domain rhetorical structures, and integration of cross-cultural discourse markers in error analysis. This refinement enhances the system's capacity to identify deviations arising from the influence of native-language discourse habits and to generate feedback that conforms to the stylistic expectations of English academic communication while maintaining consistency with professional writing norms. Feedback should be tiered according to students' writing proficiency: - Foundation Level (primarily language errors): Emphasize "error analysis + immediate correction suggestions" to build foundational language skills and satisfy competence needs. - Intermediate Level (primarily logical optimization): Incorporate "metacognitive strategy guidance" to help students master self-optimization methods and fulfill autonomy needs. - Advanced Level (primarily professional innovation): Guide "personalized academic expression" to help students develop distinctive professional writing styles, while reinforcing competence through "personalized growth recommendations." Integrating emotional feedback elements serves as a vital supplement to stimulate intrinsic motivation: Short-term encouragement should be tied to students' specific progress, avoiding vague statements—for instance, affirming concrete improvements in specialized terminology usage or logical structure optimization. Long-term growth reports must be generated periodically, presenting "writing ability growth curves" and "error reduction trends" in chart form. This allows students to visually track their progress while clarifying future improvement directions, thereby bolstering learning confidence and interest.

VI. CONCLUSION

The application of the proposed Generative AI Feedback System, grounded in Self-Determination Theory (SDT)

and Process Writing Theory, demonstrates a transformative impact on College English writing instruction, effectively shifting students from "surface-level correction" to "deep-level empowerment." The system achieves high precision in linguistic accuracy, routine error identification, and corpus-based pattern matching, while instructor verification ensures reliability for tasks requiring expert disciplinary judgment. Crucially, the "AI initial feedback + instructor refined guidance" collaborative model proved superior to traditional methods by drastically shortening the feedback cycle and offering discipline-specific guidance, which led to significant, differentiated improvements, particularly in high-level competencies like "experimental data presentation" for data-intensive majors and "linking conclusions to application prospects" for application-intensive majors.

Beyond technical skill enhancement, the system successfully stimulated student motivation, fostering a shift from extrinsic drive to intrinsic identification by fulfilling the psychological needs for autonomy (self-directed revision pace), competence (stepwise, incremental feedback), and relatedness (the humanistic touch of instructor refinement). To further optimize these outcomes, future strategies must focus on designing tiered feedback content aligned with proficiency levels and integrating emotional feedback elements, such as visual "writing ability growth curves," to sustain intrinsic motivation and guide students toward personalized academic excellence, thereby developing the clarity and precision required for technical documentation, material specification, and safety protocol writing essential in the complex manufacturing sector.

AUTHOR CONTRIBUTIONS

Lei Zhang designed, collected and analyzed the data, and drafted the manuscript. Lei Zhang conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Lei Zhang participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by, 1."An In-depth Exploration of the Impact of Feedback Systems Based on Generative Artificial Intelligence on College Students' English Writing Motivation and Ability" (Project Number: 2024WYJG028), a research project on foreign language teaching reform in ordinary undergraduate institutions in Hebei Province in 2024;

2. "Research on Cross-cultural Interpretation and Communication Strategies of Ecological Environment Discourse under the 'Dual Carbon' Background" (Project Number:

XJXM-YB-2025010), a university-level research project of Hebei University of Environmental Engineering in 2025.

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of Hebei University of Environmental Engineering, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] C. Zhou and Y. Wang, "University Students' Writing Feedback Literacy in the AI Era: The Interplay of Generative AI Acceptance, Writing Anxiety and Writing Self-Efficacy in Chinese EMI Educational Contexts," European Journal of Education, vol. 61, no. 1, Art. no. e70478-e70478, 2026, doi: 10.1111/EJED.70478.

[2] S. Jin, "Writing in the digital age: understanding the relationship between social media use and college students' writing skills and anxiety," Humanities and Social Sciences Communications, vol. 12, no. 1, pp. 1048-1048, 2025, doi: 10.1057/S41599-025-05443-1.

[3] D. Candan, "Variation in University Student Writing: A communicative text approach, Larissa Goulart, Studies in Corpus Linguistics1172024John Benjamins Publishing Amsterdam, Netherlands9789027246615ix-239.English for Specific Purposes," 2025; 79101-103, doi: 10.1016/J.ESP.2025.05.001.

[4] J. A. Hardy, "Review of Goulart (2024): Variation in University Student Writing: A Communicative Text Type Approach," International Journal of Learner Corpus Research, vol. 11, no. 2, pp. 336-341, 2025, doi: 10.1075/IJLCR.00053.HAR.

[5] N. Z. Bi and Y. Wang, "EFL College Students' Writing Self-efficacy and Strategy Use in Their Summary Writing," English Language Teaching, vol. 18, no. 3, pp. 79-79, 2025, doi: 10.5539/ELT.V18N3P79.

[6] C. Wang and Q. Li, "Effects of Gamification on Chinese EFL College Students' Writing Error Awareness and Writing Performance," International Journal of English Linguistics, vol. 15, no. 1, pp. 76-76, 2025, doi: 10.5539/IJEL.V15N1P76.

[7] Y. Zou, F. Kuek, X. Cheng, and W. Feng, "How does task complexity and task difficulty affect college students' writing performance? The mediating effect of critical thinking disposition," PloS one, vol. 20, no. 6, Art. no. e0324486, 2025, doi: 10.1371/journal.pone.0324486.

[8] M. Sinnappan and N. A. Samat, "Interlingual Error Analysis in the Writing of Tamil-Speaking College Students," Forum for Linguistic Studies, pp. 6(6), 2024, doi: 10.30564/FLS.V6I6.6949.

[9] L. Goulart, M. L. Matte, A. Mendoza, L. Alvarado, and I. Veloso, "AI or student writing? Analyzing the situational and linguistic characteristics of undergraduate student writing and AI-generated assignments," Journal of Second Language Writing, pp. 66101160-101160, 2024, doi: 10.1016/J.JSLW.2024.101160.

[10] L. Goulart and T. Dixon, "The relative influence of language backgrounds, communicative text types, and disciplines in undergraduate student writing," International Journal of Learner Corpus Research, vol. 11, no. 1, pp. 178-216, 2024, doi: 10.1075/ijlcr.23028.gou.

[11] F. Monazzah, M. Morady Moghaddam, and S. A. Ostovar-Namaghi, "The Art of Influencing: Exploring Persuasive Strategies in the Writings of Iranian University Students.Journal of Psycholinguistic Research," 2024; 53(5):62-62, doi: 10.1007/S10936-024-10106-4.

[12] L. M. Harrison, B. Challenger, E. Morgenstern, and O. A. Balarabe, "Helping College Students Write: A Guide for Educators," Taylor & Francis, 2024, doi: 10.4324/9781003402206.

[13] E. Rustan, "University Students' Plagiarism Behaviours in Writing Instruction: A Systematic Review," Journal of Language Teaching and Research, vol. 15, no. 4, pp. 1083-1094, 2024, doi: 10.17507/JLTR.1504.06.

[14] S. Ali, "A Multidimensional Analysis of Academic Writing: A Comparative Study of Saudi and British University Students' Writing," World Journal of English Language, pp. 14(2), 2024, doi: 10.5430/WJEL.V14N2P452.

[15] M. Sutherland, "Evidence of Poor Writing and Academic Standards Among University Students in the UK, and the Need for More Rigorous Accreditation of Degree Courses," The Computer Games Journal, vol. 9, no. 3, pp. 91-120, 2020, doi: 10.1007/s40869-020-00097-2.

[16] C. Han and S. Gardner, "'How can I add an argument appropriately in English?' Addition markers in Chinese L1 and English L1 university student writing," Journal of English for Academic Purposes, vol. 67, p. 101332, 2024, doi: 10.1016/J.JEAP.2023.101332.

[17] A. Sayin and N. Kahraman, "A Measurement Tool for Repeated Measurement of Assessment of University Students' Writing Skill: Development and Evaluation," Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, pp. 1-18, 2020, doi: 10.21031/epod.639148.

[18] O. Arias-Gundín and P. Robledo, "An analysis of the psychometric properties of the writing-specific cognitive strategies questionnaire for undergraduate students," Frontiers in Psychology, pp. 141274478-1274478, 2023, doi: 10.3389/FPSYG.2023.1274478.

[19] L. Dong, "Exploring the interplay between writing feedback perception and Lexical Complexity among Chinese University students: a latent Profile Analysis and Retrodictive qualitative modeling study," Reading and Writing, vol. 37, no. 10, pp. 2687-2706, 2023, doi: 10.1007/S11145-023-10489-1.

[20] Z. Miao, "The Use of Written Corrective Feedback in Writing Teaching to Chinese University Students," Journal of Humanities, Arts and Social Science, pp. 7(9), 2023, doi: 10.26855/JHASS.2023.09.007.

[21] Z. Zohoorian, "A comparative study of solitary and teacher-led pre-task planning effect on EFL university learners' writing performance: A case of an online platform amidst COVID-19," Forum for Linguistic Studies, pp. 5(1), 2023, doi: 10.18063/FLS.V5I1.1568.

[22] Q. Lu, Y. Yao, and X. Zhu, "The Effects of Writing Revisions on Writing Improvement Among Undergraduate Students: The Moderating Role of Self-rating Accuracy," The Asia-Pacific Education Researcher, vol. 33, no. 3, pp. 535-544, 2023, doi: 10.1007/S40299-023-00749-7.

[23] Q. Lu, Y. Yao, and X. Zhu, "The relationship between peer feedback features and revision sources mediated by feedback acceptance: The effect on undergraduate students' writing performance," Assessing Writing, pp. 56, 2023, doi: 10.1016/J.ASW.2023.100725.

[24] I. Ilham, B. Musthafa, and F. N. Yusuf, "UNIVERSITY STUDENTS' NEEDS OF WRITING COURSE MATERIALS: A CASE OF INDONESIA," English Review Journal of English Education, vol. 8, no. 2, pp. 31, 2020, doi: 10.25134/erjee.v8i2.2988.

[25] H. Li, "A Study on the Mechanism of College Students' Writing Therapy to Guide Rumination," Academic Journal of Humanities & Social Sciences, pp. 6(1), 2023, doi: 10.25236/AJHSS.2023.060116.

[26] S. Saleh ALYousef, "An Empirical Study of Schema Theory and Its Application in Improving University Students' Writing Ability," Arab World English Journal, vol. 11, no. 1, pp. 1-77, 2020, doi: 10.24093/awej/th.253.

[27] A. S. Alawerdy and F. S. Alalwi, "Enhancing English as a foreign language university students' writing through explicit instruction of conjunctions as cohesive devices: An experimental study," Frontiers in Psychology, pp. 131053310-1053310, 2022, doi: 10.3389/FPSYG.2022.1053310.

[28] Mujiono and S. Fatimah, "Moodle Integration Intervention in EFL Virtual Classroom and Academic Flow on University Students' Achievement in Writing," Theory and Practice in Language Studies, vol. 12, no. 10, pp. 2182-2190, 2022, doi: 10.17507/TPLS.1210.26.

[29] N. Kitajroonchai, T. Kitjaroonchai, and P. Sanitchai, "The Effects of Process Genre-Based Writing and Process Writing Approaches on Asian EFL Pre-University Students' Writing Performance," Journal of Language Teaching and Research, vol. 13, no. 4, pp. 860-871, 2022, doi: 10.17507/JLTR.1304.19.

[30] N. Justin, "Science fiction prototyping's features and impact on college students' perceptions of writing," Scientific Study of Literature, vol. 10, no. 1, pp. 128-164, 2020, doi: 10.1075/SSOL.20002.NIC.