

Research on the Application Effectiveness and Optimization Strategies of Generative AI Grading in English Writing Instruction at Tibetan Senior High Schools

Haiwan Qiu*, Tsering Deji

School of Foreign Languages and Cultures, Xizang University, Lhasa 850000, Xizang, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: This study explored Generative AI (GAI)'s applicability in English writing instruction for 53 Tibetan senior high school students in Lhasa via mixed methods, analyzing the impact of its handwritten essay grading function and identifying implementation challenges with improvement strategies. GAI performs well in basic error correction and efficient feedback, improving students' writing scores (more notably for intermediate-level and high-level learners) and stimulating their writing confidence to some extent, yet has technical limitations like recognition errors, inconsistent grading and emotionless feedback. Strategies including blended feedback, localized technical optimization and redefined teacher roles were proposed to inform digital education development in ethnic regions.

Keywords: Generative AI; English writing instruction in Xizang; Handwritten essay grading; Application effectiveness; Optimization strategies

Online publication: January 26, 2026

1. Introduction

The Outline of the National Education Development Plan (2024–2035) promotes educational digitalization for new development paths, focusing on AI-driven educational transformation^[1]. Xizang high school English writing faces hurdles from limited environments, scarce resources, and a unique trilingual context (Tibetan, Mandarin, English as first, second, third languages), resulting in students' low initial proficiency, poor vocabulary, weak grammar, and high writing anxiety^[2,3]. Traditional teacher grading is laborious, time-consuming, and infrequent (6–10 sessions/semester, 100+ essays each), failing to meet high-frequency practice needs. Generative AI (GAI) delivers rapid, real-time, comprehensive feedback on handwritten essays, addressing these pedagogical flaws^[4].

This study surveys Xizang high school students to explore GAI's application (taking Doubao as an example), identify challenges, and propose targeted optimization strategies for Xizang's plateau ethnic education context, offering a valuable micro case for advancing ethnic areas' educational governance modernization.

2. Limitations of existing research

Overseas research on Automated Writing Evaluation (AWE) has evolved from early systems like PEG and E-rater to the current GAI. Domestic studies primarily focus on the university level, with limited research on junior and senior high school stages, particularly in ethnic minority regions. Existing research exhibits two gaps as follows:

- (1) A lack of analysis on adaptability to students with weak English foundations;
- (2) Insufficient attention to technical challenges in handwritten essay recognition (as Tibetan boarding students primarily submit handwritten work).

Consequently, this study focuses on the context of weak English proficiency in ethnic minority regions, using handwritten essay recognition as an entry point to explore effective application models through empirical research.

3. Research design

3.1. Research participants

A whole-class sample of 53 Tibetan boarding students (28 males, 25 females) from Class 6, Grade 11 at Beijing Experimental Middle School in Lhasa, Xizang, was selected. Participants were grouped into three proficiency levels based on their average English application essay scores (15-point scale, averaged from four exams): Low (< 4 points, $n = 16$, 30.2%), Intermediate (4–7 points, $n = 24$, 45.3%), and High (> 7 points, $n = 13$, 24.5%). The class's final district-wide examination scores for applied writing showed no significant deviation from the regional average, ensuring its representativeness of Tibetan high schools.

3.2. Research questions

The research questions are as follows:

- (1) How does the handwritten essay grading function of generative AI impact students' writing scores, confidence, and self-directed learning abilities in Tibetan high school English composition instruction?
- (2) What challenges exist in applying this technology, such as grading accuracy and adaptability to Tibetan students' English expression habits?
- (3) How can effective optimization strategies be proposed to enhance the application effectiveness of generative AI-based handwritten essay grading in Tibetan high school English writing instruction?

3.3. Research methods and tools

A mixed-methods approach was adopted. Two 8-week rounds of action research were conducted in Grade 11's second semester, with the second round refining prompts based on student journals. After GAI graded 6 functional writing tasks, a 20-item Likert five-point questionnaire (adapted from Professor Xu Jinfen) surveyed 4 dimensions: technical efficacy, feedback quality, learning benefits, and acceptance intention^[5].

SPSS 27.0 tests showed good reliability (overall Cronbach's $\alpha = 0.907$, dimension coefficients 0.744–0.884) and validity ($KMO = 0.796$, 4 factors extracted with loadings > 0.5, cumulative variance 64.001%) (detailed metrics in **Table 1**, **Table 2**, **Table 3**). Non-normal questionnaire data was analyzed via Kruskal-Wallis test. Semi-structured interviews with 16 students (4 Low, 7 Intermediate, 5 High) and 8 teachers triangulated quantitative findings.

Table 1. Specific coefficients for each dimension of the questionnaire

Dimension	Number of items	KMO coefficient	Cronbach's Alpha coefficient
Dimension 1: Perceived technical efficacy	3	0.692	0.744
Dimension 2: Perceived feedback quality	4	0.728	0.793
Dimension 3: Perceived learning benefits	7	0.836	0.884
Dimension 4: Technology acceptance intention	6	0.822	0.845
All items across four dimensions	20	0.796	0.907

Total 2. variance explained before rotation

Component	1	2	3	4
Total	7.533	2.428	1.839	1.000
Initial Eigenvalue % of variance explained	37.666	12.141	9.194	5.000
Cumulative %	37.666	49.807	59.000	64.001
Sum of squared loadings % of variance	37.666	12.141	9.194	5.000
Cumulative %	37.666	49.807	59.000	64.001

Table 3. Component matrix of each item after rotation

Rotated Component Matrix a	1	2	3	4
Can this generative AI essay grader accurately recognize the content of your handwritten English essays?				.780
Are the English essay issues pointed out by this generative AI essay grader accurate and reasonable?				.529
Are the revision suggestions provided by this generative AI essay grader actionable?				.711
Do you find the feedback provided by this generative AI essay grader clear and easy to understand?				.718
Do you find the scoring provided by this generative AI essay grader to be objective and reliable?				.803
After using this generative AI to grade essays, do you feel your English writing skills have improved?				.690
After using this generative AI to grade essays, do you feel your interest and confidence in English writing have improved?				.574
Do you think this generative AI essay grader has been helpful for your English writing learning in the following aspects?	.705			
① Correcting basic errors such as spelling, grammar, and punctuation.				
② Optimizing overall structure, improving the introduction, conclusion, and paragraph organization of English essays.	.675			
③ Providing diverse English vocabulary and expressions to enrich language usage and avoid repetitive phrasing.	.686			
④ Organizing English writing logic to make essays more coherent and structured.	.761			
⑤ Helping you understand English writing prompts and improve your ability to analyze them.	.795			
⑥ Cultivating a unique English writing style to make your writing more distinctive.	.778			
⑦ Expand your pool of English writing materials and broaden your writing perspectives.	.640			
Do you believe this generative AI essay-grading tool could replace teachers' grading work to reduce their workload?		.542		
Do you believe this generative AI essay grader can provide personalized feedback tailored to your English proficiency and writing habits?		.830		
Do you believe this generative AI demonstrates breakthrough innovation in English essay grading technology?		.821		
Would you be willing to continue using this generative AI for essay grading in your future studies?		.572		
Would you recommend this generative AI essay grading tool to classmates or friends?		.849		
How would you rate the generative AI essay grading tool you're currently using overall?		.551		
Extraction method: Principal Component Analysis. Rotation method: Kaiser Normalized Maximum Variance method.				
a. Rotation converged after 5 iterations.				

4. Analysis and discussion of research findings

4.1. Analysis of the application effectiveness of GAI's grading function

4.1.1. Impact on students' writing skills and grades

Nonparametric tests showed significant intergroup differences in perceived writing ability improvement ($p = 0.001$) (Table 4), mainly between Low-Intermediate (Adj. $p = 0.029$) and Low-High (Adj. $p = 0.001$) groups (Table 5).

Table 4. Nonparametric test results for perceived “writing ability improvement”

Perceived improvement in English writing ability after using generative AI grading, by different English writing scores	
Independent Samples Kruskal-Wallis Test Summary	
Total N	53
Test Statistic	14.661 ^a
Degrees of Freedom	2
Progressive Significance (Two-tailed Test)	.001

a. The test statistic will be adjusted for binding values.

Table 5. Post-hoc comparison results for perceived “writing ability improvement”

Sample 1-Sample 2	Test statistic	Standard error	Standardized test statistic	Significance	Adjusted significance ^a
Low-Intermediate	11.646	4.495	2.591	.010	.029
Low-High	19.524	5.201	3.754	.000	.001
Intermediate-High	7.878	4.796	1.643	.100	.301

Applied writing scores (50 valid samples) improved: Intermediate (+2.9), High (+2.8), Low (+1.5) points (Table 6, Table 7, Table 8). Intermediate and High-level students benefited more, contradicting the “greater room for improvement in Low-level group” expectation.

Table 6. Descriptive statistics for English applied writing scores in class 6, grade 11's two final joint exams

Descriptive statistics	N	Mean	Standard deviation	Minimum	Maximum	Percentile 25th	Percentile 50th (Median)	Percentile 75th
First semester	50	2.190	1.8513	.0	9.0	.500	2.000	3.000
Second semester	50	4.620	2.5943	.0	13.0	2.750	5.000	6.000

Table 7. Nonparametric test results for English applied writing scores in class 6, grade 11, across two final joint exams

Test Statistic ^a	Second Semester English Writing - First Semester English Writing
Z	-5.252 ^b
Asymptotic Significance (Two-tailed)	.000

a. Wilcoxon signed-rank test b. Based on negative ranks.

Table 8. Changes in English applied writing scores across three ability groups in grade 11 class 6 during two final joint exams

English Writing Scores	First Semester	Second Semester
High-Level Group	4.3	7.1
Intermediate-Level Group	1.8	4.7
Low-level Group	1.0	2.5

Interview data explains this: Student 35 (Intermediate) noted, “This type of feedback helps me correct misspelled words and provides new vocabulary for reference.” Student 31 (High) stated, “It helps a lot with grammar and vocabulary because I generally have ideas for the content, but sometimes I make grammatical mistakes or don’t know which word to use.” Students with basic proficiency fully utilize GAI’s feedback.

In contrast, low-level students made modest gains. Student 1 (Low) mentioned, “I wish the corrections could be simpler, prioritizing my sentence structure. I often struggle to connect words into coherent sentences.” Their foundational language challenges exceed AI’s capacity, requiring more scaffolding support.

4.1.2. Impact on student writing interest and self-directed learning ability

According to descriptive statistics, nearly 40% of students reported that GAI grading led to an increase in their interest and confidence in writing, ranging from “some improvement” to “significant improvement” (**Table 9**). This outcome stems from the system’s instant feedback and positive reinforcement mechanisms. During interviews, students repeatedly mentioned that “feedback comes out very quickly,” “it’s very intuitive to know my weaknesses,” “I feel my efforts are rewarded,” and “when it points out the good vocabulary I used, my confidence soars.” It is precisely this positive feedback that enhances students’ learning motivation. Especially for students with weaker foundations, immediate affirmation and specific, visual improvement suggestions can reduce their writing anxiety and increase their learning confidence and interest.

Table 9. Frequency and percentage distribution of “enhanced writing interest and confidence”

After using this generative AI for essay grading, do you feel your interest and confidence in English writing have improved?	Frequency	Percentage
No improvement, even decreased	1	1.9
Little change	5	9.4
Slightly improved	26	49.1
Some improvement	18	34.0
Significant improvement	3	5.7
Total	53	100.0

In self-directed learning, students copied corrected sections, accumulated AI-recommended vocabulary, and used graded works as review resources. Student 35 (Intermediate) shifted from distrusting AI to reviewing feedback and memorizing optimized answers, showing AI promotes self-reflection and knowledge management.

4.2. Existing issues in GAI grading functionality implementation

4.2.1. Technical efficacy issues: Insufficient accuracy and stability of handwriting recognition

Although 60.4% of students believed GAI could “accurately recognize most content” or “achieve extremely high recognition accuracy,” 39.6% still encountered “unstable recognition” or worse outcomes (**Table 10**). During interviews, students complained about “unfairly losing points because crossed-out handwritten parts were still recognized, which was especially frustrating” and “higher error rates when writing in a messy hand.” Although the GAI grading system incorporates OCR recognition capabilities, it occasionally misidentifies handwritten characters like “r” and “n,” or fails to recognize crossed-out words and sentences as discarded content. This has become the primary technical obstacle affecting users’ evaluations of the system’s grading reliability.

Table 10. Frequency and percentage distribution for “accurate recognition of handwritten essay content”

Can this generative AI accurately recognize the content of your handwritten English essays?	Frequency	Percentage
Poor recognition	1	1.9
Recognition accuracy is acceptable	8	15.1
Unstable recognition	12	23.6
Most content accurately recognized	22	41.5
Extremely high recognition accuracy	10	18.9
Total	53	100.0

4.2.2. Feedback quality issues: lack of emotional warmth and personalized adaptation

Over 40% rated feedback as partially or somewhat inaccurate (**Table 11**). Students found it “robotic,” with encouragement sounding childish. Low-level students struggled with difficult vocabulary and desired sentence structure-focused correction, reflecting unique trilingual transfer challenges.

Table 11. Frequency and percentage distribution of “enhanced writing interest and confidence”

Are the English writing issues identified by this generative AI essay grader accurate and reasonable?	Frequency	Percentage
Inaccurate and unreasonable	0	0.0
Somewhat inaccurate and unreasonable	1	1.9
Partially accurate and reasonable	22	41.5
Mostly accurate and reasonable	26	49.1
Completely accurate and reasonable	4	7.5
Total	53	100.0

Regarding learning context awareness, GAI failed to account for Tibetan students’ “memorize first, write later” learning habits (both teachers and students stated in interviews that “students memorize model essays before writing content”). This resulted in feedback recommendations disconnected from students’ existing knowledge frameworks, and scoring criteria that inadequately reflected their characteristic of “weak foundations but capable of writing basic essays.” Students interviewed suggested adjusting scoring weights, such as reducing the proportion of deductions for basic errors (spelling, grammar) and increasing the proportion for content completeness (aligning with topic requirements and covering core content).

4.2.3. Inadequate originality detection and guidance mechanisms

GAI poorly detects plagiarism, where Student 1 (Low) noted copied model essays are scored well, encouraging opportunistic behavior. Unlike teachers who balance encouragement and critique, GAI lacks human intervention to guide independent thinking.

4.2.4. Teacher acceptance issues: cumbersome operational processes and concerns over scoring standards

Teachers widely acknowledge GAI's strengths in providing "comprehensive and meticulous" grading, but "excessive operational complexity" (such as individually photographing and uploading assignments for recognition) remains the primary obstacle hindering its adoption. One teacher bluntly stated: "The prep work is too much, I'd rather just grade it myself" and "I'd rather grade it myself for peace of mind." Additionally, teachers expressed concerns about "inconsistent grading standards," noting that "some students with similar content receive vastly different scores," which undermines trust in GAI's scoring among both teachers and students.

5. Optimization strategies

5.1. Establishing a "human-machine collaboration with teacher guidance" teaching model

GAI handles initial grading (spelling, grammar errors), while teachers address common issues collectively and provide individualized guidance, focusing on low-level students. Teachers should calibrate GAI scores (e.g., reducing basic error penalties for low-level students) to avoid undermining confidence, as students find GAI overly strict.

To curb mechanical copying, teachers design multi-revision tasks and group activities. Student 34 (Intermediate) admitted mechanical copying; guiding active reflection on feedback cultivates metacognition. One student suggested in-class writing without references before GAI correction to reflect real ability and deepen impressions.

5.2. Advancing technology localization and adaptive optimization

GAI should upgrade handwriting recognition to ignore crossed-out content. A trilingual (Tibetan-Chinese-English) analysis mechanism reduces negative transfer. Uploading student background data (proficiency, grades) enables adaptive feedback: basic support for Low-level students, fluency and depth guidance for High-level ones, aligning with their zone of proximal development.

5.3. Streamlining workflows and redefining teacher roles

To enhance teachers' acceptance and utilization of GAI for grading handwritten English essays, overcoming facility barriers is crucial. Schools must proactively facilitate technical integration and equipment procurement (e.g., dedicated scanners), design bulk upload and one-click grading functions within common teaching platforms, such as standardizing writing paper for easier scanning. After resolving hardware issues, organize training to guide teachers on effectively inputting prompts (e.g., controlling feedback vocabulary difficulty to avoid exceeding student comprehension), designing tiered teaching strategies based on AI report, identified needs for different proficiency levels, and redirecting saved time toward more valuable teacher-student interaction and instructional design. Teachers must also recognize that AI serves as a "teaching assistant" rather than a replacement. They should act as interpreters, supplementers, and coordinators of AI feedback, not merely technical operators.

6. Conclusions and future directions

GAI's handwritten grading positively enhances Tibetan high school students' writing performance and interest, but faces challenges in recognition accuracy, personalization, and teacher acceptance. Future efforts should build a student-centered

“human-machine symbiosis” ecosystem, with AI handling common tasks and teachers focusing on individual needs, especially supporting low-proficiency students. Limitations include a single-school sample and short research period. Future studies should expand samples, conduct longitudinal tracking, and explore GAI’s application in other language skills (listening, speaking, reading) in Xizang.

Funding

Xizang University Graduate High-Level Talent Cultivation Program (Project No.: 2025-GSP-S108)

Disclosure statement

The author declares no conflict of interest.

References

- [1] The Central Committee of the Communist Party of China and the State Council, 2025, The Outline of the Plan for Building an Education Powerhouse (2024–2035). *People’s Daily*, January 20, 2025, 6.
- [2] Yangjin Q, 2011, Research on the Current Status and Countermeasures of English Teaching in Tibetan Primary and Secondary Schools. *Journal of Xizang University (Social Sciences Edition)*, 26(3): 163–169.
- [3] Xiao C, Chen H, 2023, Psychological Distress and Coping Strategies for Tibetan Students’ Lack of Motivation in English Learning. *Journal of Xizang University (Social Sciences Edition)*, 38(3): 233–240.
- [4] Liu M, Zou W, Zong M, et al., 2025, How Generative AI Empowers Teaching: An Empirical Study on English Writing Instruction for Secondary School Students. *China Educational Technology*, 2025(6): 70–79.
- [5] Zhang S, Xu J, 2019, The Impact of Online Automated Feedback on Writing Among English Learners of Different Proficiency Levels from a ZPD Perspective. *Foreign Languages and Foreign Language Teaching*, 2019(5): 30–39.
- [6] Fang F, Wu Q, 2025, Exploring Metalinguistic Practices and Learner Agency in English Media Instruction under Generative AI. *Frontiers in Foreign Language Education Research*, 8(4): 13–23.
- [7] Sun P, Xu J, Zhang J, 2025, Research on the Value-Added Benefits of Generative AI Feedback for College Students’ English Writing Quality. *Frontiers in Foreign Language Education Research*, 8(4): 24–36.
- [8] Lu Y, Tang X, 2025, Hierarchical Levels and Progression Pathways of Generative AI-Empowered Classroom Instruction. *Research on Educational Technology*, 46(6): 75–82.
- [9] Li M, Yang Y, 2025, Rational Understanding of the Promoting Role of Educational Digitalization in Educational Transformation. *Frontiers of Foreign Language Education Research*, 8(4): 37–44.

Publisher’s note

Whioce Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.