Available Online: 4 November 2026 DOI: 10.54254/2753-7102/2026.38617

Generative AI-enabled writing feedback in primary school English instruction: application landscape, practical dilemmas, and a teacher-led collaborative framework

Lunyuan Wang

Jiaying University, Meizhou, China

iclara1314@gmail.com

Abstract. Writing feedback is a pivotal yet chronically constrained component of primary school English writing instruction, squeezed by large class sizes, low marking frequency, and homogeneous comments. Generative Artificial Intelligence (GenAI), with its immediacy, personalization, and dialogic interaction, opens new possibilities for reconstructing writing feedback. Drawing on a narrative synthesis of 52 Chinese and international sources, and addressing a literature in which research on AI writing feedback concentrates on higher education and on the learner side, this paper maps the application landscape of GenAI-enabled writing feedback in primary school English, consolidates and examines five practical dilemmas concerning accuracy, appropriateness, teacher literacy, ethics, and evaluation orientation, and proposes a Teacher-led, AI-assisted, Student-centered Collaborative Feedback Framework (TAC-F). The framework assigns teachers responsibility for goal setting, affective support, and feedback verification, positions GenAI as an initial assessor and material provider, and treats students as reflective revisers within a pre-writing, while-writing, and post-writing closed loop, supported by prompt-template libraries, feedback-verification routines, school-based teacher AI literacy development, and ethical data norms. Preliminary multi-tool observations, together with a six-dimension observation framework and a same-prompt pilot protocol, are developed to help teachers screen AI-generated feedback; the observations suggest that current tools are reliable on explicit errors but unstable at the discourse level. Six implementation suggestions are advanced to inform the digital transformation of primary school English writing instruction.

Keywords: generative artificial intelligence, primary school English, writing feedback, human-AI collaboration, teacher digital literacy

1. Introduction

Writing is a core component of language ability in the English curriculum and a concentrated expression of the quality of joyful and effective learning. *The English Language Curriculum Standards for Compulsory Education (2022 Edition)* [1] establish core competencies as the organizing principle and require expressive ability to be developed through authentic language practice. Writing feedback, as the hinge connecting writing and improving writing, largely determines the effectiveness of writing instruction. At the primary level,

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

however, feedback has long been trapped in a state of high demand and low supply. In commonly reported practice, a single English teacher teaches four to six classes and is responsible for the writing of two to three hundred students, which makes detailed marking of every essay infeasible. Feedback concentrates on static error correction of spelling and grammar, lags behind the writing process, and rarely offers differentiated suggestions for individual differences. Students who receive marked compositions often look only at the grade without revising the problems. The low frequency, delay, and homogeneity of feedback make the ideal of assessment for writing difficult to realize.

Two developments are changing this picture. On the policy side, in November 2024 the General Office of the Ministry of Education issued *the Notice on Strengthening Artificial Intelligence Education in Primary and Secondary Schools* [2], which calls for the basic popularization of AI education in primary and secondary schools before 2030; *the National Education Strengthening Plan (2024-2035)* [3] further deploys the national strategy of education digitalization and the use of AI to support educational change. On the technology side, Generative Artificial Intelligence (GenAI) built on Large Language Models (LLMs) has shown general-purpose text understanding and generation capabilities [4]. Researchers have systematically discussed the overall impact of ChatGPT on foreign language education [5, 6], reviewed the historical trajectory and future pathways of AI education in Chinese primary and secondary schools [7], examined the practical logic of integrating GenAI into the national smart education platform [8], and repeatedly reminded the field that technological empowerment must return to authentic teaching needs [9]; Jiao was among the first scholars in China to call public attention to the impact of ChatGPT on schooling [10]. Empirical evidence on GenAI writing feedback in secondary and tertiary EFL settings is accumulating rapidly [11, 12], but existing reviews commonly note that the literature displays a pattern of "strong in higher education, weak in basic education, rich in ideas, poor in implementation" [13].

The synthesis underlying this paper is narrative rather than systematic: sources were selected for topical relevance and journal quality, mainly SSCI- or CSSCI-indexed venues, complemented by four policy and standards documents. The gap claims below are therefore conclusions from this reviewed corpus rather than exhaustive statements about the literature. Within that corpus, three gaps stand out at the intersection of primary school English writing feedback. First, a school-level gap: empirical studies of AI writing feedback concentrate on university EFL learners [11, 12, 14]; a small number of studies touch primary contexts through creative writing and oral practice, and reports on written feedback are limited to senior secondary and lower-secondary contexts [15, 16]; to our knowledge, the reviewed sources contain no dedicated study of feedback mechanisms for primary school writing. Second, an agent gap: existing studies almost always examine feedback uptake and engagement from the learner's side [11, 14]; although teacher-side engagement with automated feedback has begun to attract international attention, how primary school teachers, as the first persons responsible for feedback, use, verify, and transform AI-generated feedback remains, to our knowledge, unexamined in Chinese-language core journals. Third, a method-and-scenario gap: discussions in China are dominated by macro-level argumentation, classroom-based operational studies of specific subjects are scarce, and qualitative methods such as interviews and classroom observation are particularly rare. Accordingly, this study poses three research questions. RQ1: What application landscape does GenAI open up for primary school English writing feedback? RQ2: What practical dilemmas does its implementation face? RQ3: How can a teacher-led improvement framework and implementation pathway be constructed?

2. Literature review and theoretical basis

2.1. From automated writing evaluation to generative feedback

The technological history of writing feedback has moved from Automated Writing Evaluation (AWE) to generative feedback. AWE typically relies on manually constructed scoring rules and feature templates and is strong at scores and static comments, although later work has integrated automated feedback into process-oriented writing instruction [16]. Research in China on automated feedback for college English writing accelerated in the mid-to-late 2010s [17], and experience in intelligent marking accumulated in large-scale speaking examinations [18]. Its limits are equally clear: restricted feedback dimensions, rigid language adaptation, and little capacity to respond to students' specific problems. GenAI changes this picture by trading part of the reliability and calibration of rule-based scoring for breadth and dialogicity: rather than scoring by rules, it understands the text and converses, providing contextualized annotations and revision examples concerning content, organization, and language. Comparative experiments have shown that students receiving ChatGPT feedback outperformed those receiving traditional AWE feedback in both writing performance and ideal L2 writing self [12]. Students adopt teacher feedback and ChatGPT feedback in systematically different ways [11]; the two kinds of feedback also differ in textual features and in the patterns of student engagement they activate [19], suggesting that human and AI-generated feedback may complement rather than replace each other [14]. Research in China has begun to rebuild classroom evaluation schemes with GenAI [20], and empirical work on LLM-empowered oral English teaching has confirmed the feasibility of LLMs in language training scenarios [21]. This paradigm shift provides the technical possibility of an AI initial assessment, teacher transformation, and student revision sequence for primary school English writing feedback.

2.2. GenAI and foreign language education: three lines of domestic research

Research in China on GenAI and foreign language education has unfolded along three lines. The first is theory building. Wen and Liang proposed human-machine interactive negotiation competence, defined as the learner's capacity to negotiate meaning with intelligent agents and to accomplish tasks collaboratively [22]; Yang reflected, from the perspective of second language acquisition, on how LLMs reshape the input-interaction-output chain [23]; Zheng systematically reviewed the key controversies of GenAI in foreign language education, arguing that debates such as replacement or empowerment, and equity or divide, are ultimately about the ordering of educational values [24]. The second is the application landscape. Hu and Qi offered the master frame of seeking change and responding to change [5]; Xu and Zhao defined the multiple roles of LLMs in English teaching as resource, tool, and co-learner [25]; Qin and Lu classified application modes along the line of language competence [26]; Hong et al. clarified the internal logic and practical pathways of GenAI-empowered foreign language education [27]; Xu constructed a teaching and research framework for AI-empowered cognitive development in English reading and writing [28]; and Xu further systematized implementation strategies for personalized foreign language education in the context of LLMs [29]. The third is teacher development. Wang and Zhou constructed the connotation and framework of digital-intelligence literacy for university foreign language teachers [30]; Zhang explored pathways for integrating LLMs into foreign language teacher education [31]; and the survey by Cui and Schunn revealed a structural gap between teachers' and students' AI literacy [32]. Overall, all three lines take higher education as the default scenario; primary school English appears only in passing, and no dedicated study appears in the reviewed corpus. This is the direct starting point of the present study.

2.3. Teacher-side human-AI collaboration

International research has likewise moved from AI replacing teachers to human-AI complementarity. Tai and Chen's quasi-experiment showed that GenAI chatbots can effectively support upper-primary students' oral English practice, with paired interaction outperforming individual interaction [33]; Jeon's qualitative study of young learners mapped the affordance boundaries of classroom AI chatbots [34]; Yang's professional development program demonstrated that intervention can significantly raise teachers' self-efficacy for integrating AI into English speaking teaching [35]. Meanwhile, affordance-based comparisons of the effectiveness of human and intelligent tutors [36] and critical analyses of the triple illusions of AI-empowered teacher development [37] jointly remind us that the condition for technological empowerment is the redefinition, not the dissolution, of the teacher's role. Modeling of teacher-student-machine cooperative games in China based on symbiosis theory [38], together with the affordance perspective on tutor effectiveness [36], provides a direct theoretical precursor for the triadic collaboration constructed in this study.

2.4. Theoretical foundations

Three bodies of theory underpin this study. First, formative feedback theory: the value of feedback lies in narrowing the gap between the current state and the goal, and its effectiveness depends on comprehensibility, actionability, and affective safety; this provides the criteria for judging the quality of AI-generated feedback. Second, human-AI collaborative teaching theory: teachers and AI occupy different ecological niches in the instructional system, and the key to collaboration lies in task allocation and responsibility attribution. AI undertakes the automatable functions of initial assessment and provision, while teachers retain goal decision, value guidance, and final adjudication. Third, the zone of proximal development: appropriate feedback should fall within the learner's reachable-with-effort zone, which imposes a hard constraint on the language-level adaptation of feedback at the primary level, precisely the weakest link of current AI-generated feedback.

3. The application landscape: how GenAI can reconstruct writing feedback

3.1. Three transformations in feedback provision

Compared with teacher marking and AWE, GenAI brings three transformations on the supply side of writing feedback. The first is immediacy: annotations return as soon as an essay is submitted, compressing the feedback cycle from weeks to minutes and making immediate revision possible while the draft is still fresh. The second is personalization: analyses with varying granularity and wording can be generated for students of different levels in the same class, offering a scalable solution for differentiated feedback, consistent with the implementation logic of personalized foreign language education in the context of LLMs [29]. The third is dialogicity: students, or teachers acting on their behalf, can follow up on a comment and request examples, turning feedback from a one-off verdict into multi-turn negotiation, which is precisely the cultivation ground for the human-machine interactive negotiation competence proposed by Wen and Liang [22]. Preliminary international evidence from primary contexts also supports feasibility: GenAI-supported creative writing improved primary students' writing performance, self-efficacy, and flow experience simultaneously [39].

3.2. A six-dimension observation framework and a multi-tool pilot protocol

To obtain first-hand evidence, this study designed a multi-tool observation protocol in which one and the same essay was submitted to all tools under a single unified prompt. The material is a Grade-5 (second semester) essay containing nine planted errors, comprising five tense, two spelling, and two collocation errors. The same

teacher marking instruction was sent to three mainstream tools (Doubao, iFLYTEK Spark, and ChatGPT), and their outputs were compared along six dimensions: error capture rate, dimensional completeness, feedback granularity, comprehensibility, hallucination risk, and encouragement alignment. The protocol and the essay are described in Appendix A. The ethical stance of this observation is teachers using AI to generate marking references: students never directly use GenAI, and no personal information of minors is collected. Table 1 presents the six-dimension framework, which this study proposes as an operational tool for teachers to screen AI-generated feedback.

Table 1. A six-dimension observation framework for AI-generated writing feedback

Dimension	Guiding question	What to record
Error capture rate	Of the nine planted errors (five tense, two spelling, two collocation), how many does the tool identify?	Count and type of detected and missed errors
Dimensional completeness	Does the feedback cover language, content and structure, and conventions?	Coverage across the three categories
Feedback granularity	Does the tool quote the original sentence and give a concrete revision?	Number of sentence-level annotations
Comprehensibility	Do the suggested vocabulary and grammatical terminology stay within the word lists and level of the national curriculum standards?	Out-of-level vocabulary and terminology instances
Hallucination risk	Does the tool alter expressions that were originally correct?	Each altered-correct instance with quotation
Encouragement alignment	Does the tone of the comments match the severity of the problems?	Tone-problem consistency notes

Note. Level 2 of the national curriculum standards denotes the proficiency expected by the end of Grade 6 under *the English Language Curriculum Standards for Compulsory Education (2022 Edition)* [1]; its annexed word lists serve as the judging benchmark for comprehensibility.

Preliminary observation across the three tools yields three qualitative findings. First, all tools produce structurally complete and friendly-toned feedback and reliably capture explicit spelling and tense errors; capture of the two collocation errors was not analyzed separately in this informal pass. Second, performance is inconsistent in three deeper capabilities: judgments of discourse-level tense consistency, such as maintaining past tense across a narrative, are unreliable; the tools lack age calibration on whether an error should be corrected and to what extent, and suggested vocabulary often exceeds Level 2 of the national curriculum standards, that is, the proficiency expected by the end of Grade 6; and isolated tools occasionally correct originally right expressions into wrong ones, a hallucination phenomenon. These findings converge with international conclusions that the specificity advantages and the accuracy risks of AI-generated feedback coexist [11, 40]. It must be emphasized that the observations reported here are informal and illustrative rather than a systematic evaluation; they derive from single-pass use of the three tools by the author, and Appendix A specifies the formal two-coder protocol that a documented replication should follow. The coding scheme and an archive of raw outputs are available from the authors and provide a documented basis for a future systematic evaluation.

3.3. Application modes across the writing process

Combining the literature with the pilot observation, the application of GenAI in primary school English writing feedback can be summarized in three modes. In pre-writing, GenAI generates topic scaffolds, sentence frames, and model texts according to the writing task, and the teacher screens them for task fit before distribution, lowering the entry threshold of not knowing where to start. In while-writing, GenAI serves as a classroom assistant to the teacher during face-to-face conferencing, for instance by instantly checking idiomatic expressions or generating contrastive examples, while the teacher retains the lead in explanation. In post-writing, GenAI performs an initial assessment of the whole class's essays and produces marking references; the teacher verifies, filters, and rewrites them before distribution, and students revise accordingly. This division echoes the hierarchical modes of GenAI classroom application proposed by Lu and Tang [41] and corroborates the language-competence-based classification of Qin and Lu [26]; explorations of AI oral-practice agents built on textbook characters [42] show that subject-specific agent design also has realistic space in primary contexts. The common feature is that GenAI always occupies the assistant-provision position, and the final form of feedback is decided by the teacher.

4. Practical dilemmas: five tensions behind technological empowerment

The five dilemmas below are consolidations of issues documented mainly in higher-education and general AI-in-education contexts, examined here for their specific manifestations in primary school English writing feedback.

4.1. The accuracy dilemma: incomplete error capture and hallucinated corrections

The quality of GenAI feedback is reliably accurate on explicit errors but unstable on deeper capabilities. The pilot observation suggested reliable capture of explicit spelling and tense errors but inconsistent performance on discourse-level judgments such as maintaining past tense across a narrative. More alarming is hallucinated correction, in which the tool judges a student's correct expression to be wrong and offers a wrong model. Research on bias in intelligent learning systems warns that personalized systems carry bias risks in which systematic and random errors coexist [43]; international comparative studies likewise find that ChatGPT feedback and direct human corrective feedback each have strengths and weaknesses in accuracy [40]. For primary school English teachers this means AI-generated feedback cannot be forwarded as-is. If teachers lack the ability to recognize misjudgments, the machine's confident errors acquire authoritative endorsement through written comments, doing more harm than giving no feedback at all.

4.2. The appropriateness dilemma: mismatch between feedback wording and school level

Formative feedback theory requires feedback to fall within the learner's zone of proximal development, whereas GenAI by default produces feedback in the register addressed to mature writers. In the pilot observation, revision suggestions frequently used vocabulary beyond Level 2 of the curriculum standards, and grammatical terms such as past participle and attributive clause exceeded primary students' metalinguistic capacity. Xu and Zhao's definition of LLM teaching roles suggests that AI feedback needs role-based reconstruction, output in the voice of a co-learner rather than a judge [25]; international research also shows that feedback effects depend on how learners process it [14]. Analyses of the constraints on integrating LLMs into primary school Chinese teaching likewise indicate that school-level adaptation is the first threshold that intelligent technology must cross to land in a subject [44]. Without school-level transformation, the very professionalism of AI-generated feedback becomes a comprehension barrier for students.

4.3. The literacy dilemma: the structural gap in teacher AI literacy

Teachers are the gatekeepers of AI-generated feedback, and their literacy determines how far the technology dividend can be realized. Zheng et al. criticized the triple illusions of AI-empowered teacher development, pointing out that tool availability does not equal capability availability [37]. Although the Ministry of Education standard Teacher Digital Literacy (JY/T 0646-2022) [45] establishes a five-dimension framework, to our knowledge no standard yet specifies the subject-specific capability of AI feedback verification in English teaching. Wang and Zhou's framework research [30] and Cui and Schunn's empirical survey [32], both conducted in higher education, suggest that teachers' AI literacy may remain at a shallow state of being able to use but not to audit, and daring to use but not to trust. Large-scale surveys of primary and secondary school teachers similarly reveal an obvious gap between the breadth and the depth of application [46]: most teachers can complete basic dialogues but lack the ability to constrain outputs through prompt design or to identify hallucinations through cross-validation. International experience shows that this gap can be narrowed by professional development interventions [35], but the interventions themselves need school-based carriers and institutional support.

4.4. The ethics dilemma: data security, evaluation equity, and the protection of minors

ChatGPT's educational application has been described from the start as a double entity of opportunity and risk, Aladdin's lamp and Pandora's box [47]. Entering the scenario of primary school writing feedback, its ethical tensions are especially sharp and naturally touch three critical boundaries. The first is data security. Essay texts are minors' personal information, and bulk uploading them to third-party large models carries compliance risks; de-identification and localized deployment must be the baseline [48]. The second is evaluation equity. Differences in family digital conditions may translate into differences in students' access to AI feedback, with the risk of stratified transmission along the digital divide; Wang and Kou's analysis of the triple structure of the digital divide reminds us that the access, usage, and literacy divides are superimposed and amplified inside and outside the classroom [49]. The third is evaluation ethics. AI-generated feedback carries the authoritative illusion of machine scoring; if teachers adopt it without verification, they not only weaken the professionalism of evaluation but also send students the wrong signal that the machine outranks the teacher. Research on ethical risk management of GenAI in education [48] and implementation explorations of AI ethics education [50] both stress that ethical governance must be embedded in the application process rather than appended as after-the-fact review.

4.5. The orientation dilemma: the risk of feedback degenerating into an error hunter

The most hidden dilemma lies in evaluation orientation. The function of writing feedback is not only error correction but also arousing the willingness to revise, cultivating audience awareness, and building expressive confidence. GenAI is naturally good at finding faults; if used improperly, it can easily drag writing instruction back to the error-list mode, in which students busy themselves with item-by-item correction yet never experience rewriting for a real reader. Xu's research on human-AI collaborative writing engagement shows that the realization of feedback value depends on students' deep investment in behavior, cognition, and emotion [51]; Yan's experiment further showed that the gain of ChatGPT feedback for writing improvement is significantly higher when it is processed through teacher and peer collaboration than when it is handled individually [14]. In other words, AI-generated feedback avoids an alienated outcome, in which efficiency rises while educational substance is diluted, only when it is embedded in the educational relations of teacher-student interaction.

5. The improvement framework: Teacher-led, AI-assisted, Student-centered Collaborative Feedback (TAC-F)

5.1. Overall structure

In response to these dilemmas, and drawing on symbiosis-theoretic modeling of teacher-student-machine relations [38] and affordance analysis of human-AI collaboration [36], this study constructs the Teacher-led, AI-assisted, Student-centered Collaborative Feedback framework, abbreviated TAC-F, whose structure is shown in Figure 1. The abbreviation is partial, compressing Teacher, AI, and Collaborative Feedback; student-centeredness is carried in the full name. The framework answers three questions: who does what, through role division among the three agents; how, through the three-stage process loop; and on what conditions, through the four guarantee mechanisms. Relative to its precursors, the increment of TAC-F is threefold, and is stated here to delimit what is transferred and what is new. The triadic role division transfers the teacher-student-machine modeling of symbiosis theory [38] into the writing feedback scenario; the three-stage loop adapts the hierarchical modes of GenAI classroom application [41] to the rhythm of composition teaching. What is new, beyond this domain instantiation, is, first, the elevation of feedback verification from a professional exhortation to a rigid institutional routine, in which AI initial assessments may not reach students without teacher verification and records of teachers' rewriting are retained for review; second, the six-dimension observation framework of Table 1, an operational instrument for screening AI-generated feedback at a given school level; and third, the genre-by-level-by-focus prompt template library, which turns school-level adaptation into a reusable school asset.

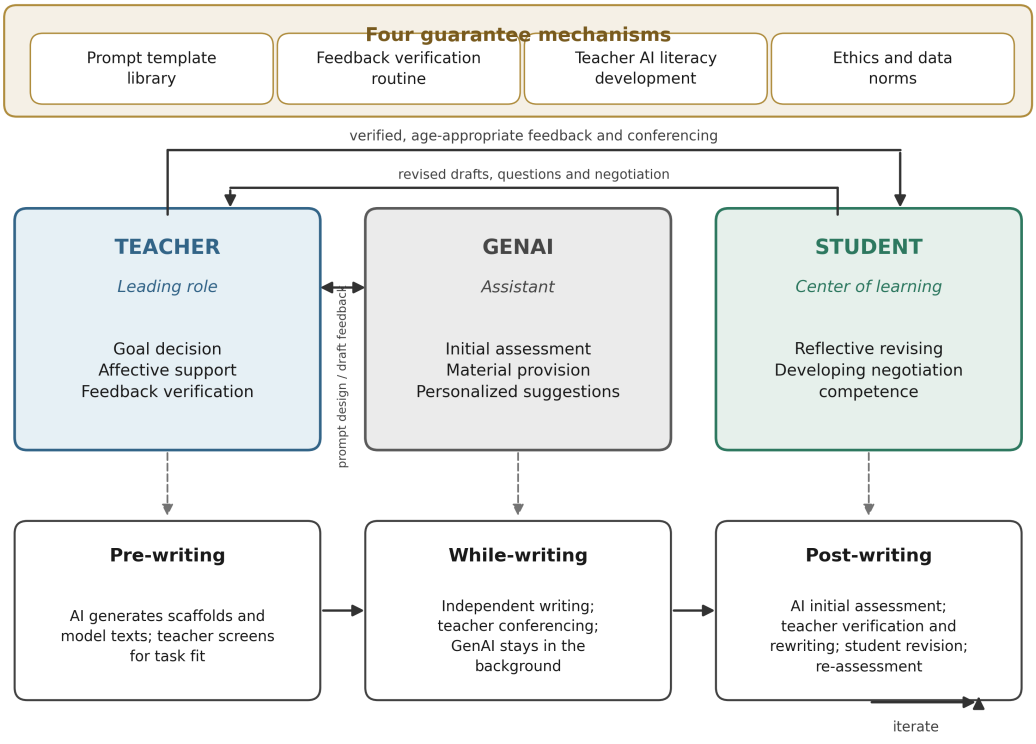


Figure 1. Overall structure of the TAC-F framework

5.2. Role division among the three agents

The teacher is the leader and the ultimately accountable party of the feedback system, undertaking three kinds of duties: goal decision, determining writing tasks and feedback focus according to the curriculum standards and learner conditions; affective support, maintaining writing motivation through conferencing and comments; and feedback verification, screening, correcting, and rewriting AI initial assessments. GenAI is the assistant provider, undertaking initial assessment, generating marking references under unified prompts; material provision, supplying scaffolds, model texts, and contrastive examples; and personalized suggestions, offering practice plans for students of different levels. The student is the agent of reflection and revision, whose core task is to complete revision and improvement based on comprehensible feedback and to develop negotiation competence in human-machine interaction [22]. This division implements student-centeredness as a structural position in the framework. The responsibility boundaries are clear: AI output is only a reference document, the pedagogical transformation of feedback is completed by the teacher, and the ultimate agent of revision is the student.

5.3. The three-stage closed loop

In the pre-writing stage, GenAI generates writing scaffolds and model texts, and the teacher screens them for task fit before release, addressing students' difficulty in getting started. In the while-writing stage, students write independently while the teacher circulates among them and holds face-to-face conferences, preserving the authenticity of the writing process and the presence of teacher-student interaction. In the post-writing stage, GenAI performs initial assessment to generate marking references; the teacher verifies and transforms them into age-appropriate comments; students revise and improve accordingly; and the revised drafts re-enter AI initial assessment, forming the iterative loop of initial assessment, verification, revision, and re-assessment. The design intent of the loop is to anchor the comparative advantages of AI, namely speed, coverage, and consistency, at the initial-assessment node; the irreplaceable functions of the teacher, namely judgment, transformation, and motivation, at the verification node; and the mechanism of learning, namely revision practice, at the student node.

5.4. Four guarantee mechanisms

The effective operation of the framework depends on four institutional supports. The first is the prompt template library: reusable prompt templates prefabricated by genre, level, and feedback focus constrain AI output within the school-level register and lower the operational threshold for individual teachers. The second is the feedback verification routine: a rigid process stipulating that AI initial assessments may not reach students without teacher verification, with records of teachers' revisions retained for teaching-research review. The third is school-based teacher AI literacy development: taking the Teacher Digital Literacy standard [45] as the outline, case-based training is organized around the three special capabilities of prompt engineering, hallucination recognition, and feedback transformation, drawing on the four-dimension structure of affective, behavioral, cognitive, and ethical AI literacy [52]. The fourth is the ethics and data security norm: de-identification before essay upload, preference for locally deployed tools, and explicit boundaries for parental informed consent and data retention [48]. In this mapping, the prompt template library and the feedback verification routine jointly answer the accuracy and appropriateness dilemmas; school-based development answers the literacy dilemma; the ethics norm answers the ethics dilemma; and all four serve the reorientation of evaluation purpose that guards against the orientation dilemma, making the framework executable as a whole.

6. Implementation suggestions

Corresponding to the five dilemmas, six operational suggestions are proposed. First, building the library: with school-based teaching-research groups, the standard collegial professional-development unit in Chinese schools, as the unit, compile prompt template libraries along the three dimensions of genre, level, and feedback focus, and iterate each semester on observational data; this addresses the appropriateness dilemma at its source and scaffolds the capability that the literacy dilemma demands. Second, setting the rule: write AI initial assessment, teacher verification, and student revision into school homework management regulations, specifying the accountable party and the time limit of each link. This institutionalizes the verification routine that answers the accuracy dilemma and, by preventing AI-generated feedback from bypassing teachers and reaching students directly, also answers the orientation dilemma. Third, developing staff: bring AI feedback verification into school-based training and teaching-research activities, using case-review training with real essays and real AI outputs, and guarding against training for its own sake in light of the triple illusions [37]; this implements the resolution of the literacy dilemma. Fourth, coordinating home, school, and community: safeguard equity of use by opening school computer rooms and after-school service periods to students lacking devices, blocking the transmission of the digital divide into feedback opportunity [49]. Fifth, holding the line: establish operational rules of de-identification, minimization, and time-limited deletion for essay data, prefer locally deployed models that have completed regulatory registration for education-sector use, and hold the baseline of the ethics dilemma [48]. Sixth, repositioning: evaluate feedback effectiveness by writing competence rather than the number of corrections, bringing revision quality, progress in second drafts, and writing self-efficacy into observation, so that technology always serves the core-competency orientation of the curriculum standards [1].

7. Conclusion and outlook

This study responds to the anxiety that "the landscape is clear but the path is blurred" when GenAI enters English instruction in basic education. Through a narrative synthesis of 52 Chinese and international sources, 39 in Chinese and 13 in English, together with a multi-tool same-prompt pilot observation, it has mapped the application landscape of GenAI-enabled primary school English writing feedback, consolidated and examined five dilemmas of accuracy, appropriateness, literacy, ethics, and orientation, and constructed the operational TAC-F framework together with six implementation suggestions. Sections 3, 4, and 5 answer RQ1, RQ2, and RQ3 in turn.

The framework's logic is not tied to the Chinese context. What travels is the division of labor in which AI drafts and teachers verify, the verification routine with its audit trail, and the six screening dimensions; what requires local substitution are the curriculum-level anchors, the teacher literacy standard, and the specific toolset. Systems with comparable class sizes and marking loads could adapt TAC-F by substituting their own level benchmarks for Level 2 and their own governance documents for the standards cited here.

The boundaries of this study must be acknowledged. The effectiveness of the framework has not yet been tested through classroom experiments, and the pilot observation covers a single essay and three tools in an informal mode; the conclusions are advisory rather than causal. These boundaries mark the agenda for subsequent research: first, qualitative research on primary school English teachers' adoption of AI feedback, revealing the micro-mechanisms of verification and transformation behavior; second, a quasi-experimental comparison of AI initial assessment plus teacher verification against teacher-only marking, testing the net effect of the framework on writing scores, revision quality, and writing self-efficacy; third, development and validity verification of a prompt template library for primary school English writing feedback; fourth,

conceptual reconstruction and measurement of teacher feedback literacy in human-AI collaboration scenarios. Writing feedback is the teaching activity that technology can least replace and that most depends on educational judgment. Let AI do what it is good at, let the teacher return to what is uniquely human, and let the student truly learn to write between the two. This is the entire intent of the TAC-F framework.

References

[1] Ministry of Education of the People's Republic of China. (2022). *English language curriculum standards for compulsory education (2022 edition)*. Beijing Normal University Press.

[2] General Office of the Ministry of Education. (2024, December 2). *Notice on strengthening artificial intelligence education in primary and secondary schools (Jiao Ji Ting Han [2024] No. 32)*. http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/202412/t20241202_1165500.html

[3] Central Committee of the Communist Party of China & State Council. (2025, January 19). *National education strengthening plan (2024–2035)*. https://www.gov.cn/gongbao/2025/issue_11846/202502/content_7002799.html

[4] Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

[5] Hu, J. S., & Qi, Y. J. (2023). China's foreign language education in the ChatGPT era: Seeking change and responding to change. *Technology Enhanced Foreign Language Education*, (1), 3–6.

[6] Zhang, Z. Y., & Hong, H. Q. (2023). ChatGPT-supported foreign language teaching: Empowerment, problems, and strategies. *Foreign Language World*, (2), 38–44.

[7] Wu, D., Li, H., & Wu, H. J. (2026). A historical review and future pathways of artificial intelligence education in China's primary and secondary schools. *Education Research, 47*(6), 147–159.

[8] Sun, L. H., & Zhou, L. (2024). The practical logic of integrating generative artificial intelligence into the National Smart Education Platform for Primary and Secondary Schools. *China Educational Technology*, (8), 71–79.

[9] Gu, X. Q., Wang, C. L., Wang, P. J., & Jing, Y. H. (2025). Mechanisms, demands and pathways of generative artificial intelligence empowering teaching. *Journal of the Chinese Society of Education*, (4), 15–22.

[10] Jiao, J. L. (2023). ChatGPT: Friend or foe of school education? *Modern Educational Technology, 33*(4), 5–15.

[11] Zou, S., Guo, K., Wang, J., & Liu, Y. (2026). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. *Computer Assisted Language Learning, 39*(4), 1062–1091. https://doi.org/10.1080/09588221.2024.2447279

[12] Shi, H., Chai, C. S., Zhou, S., & Aubrey, S. (2026). Comparing the effects of ChatGPT and automated writing evaluation on students' writing and ideal L2 writing self. *Computer Assisted Language Learning, 39*(4), 1148–1175. https://doi.org/10.1080/09588221.2025.2454541

[13] Sun, C. Z., & Liu, W. Y. (2025). Artificial intelligence-empowered foreign language teaching: Applications and future prospects. *Foreign Languages and Their Teaching*, (5), 77–87.

[14] Yan, D. (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology, 28*(1), 1–16.

[15] Ozan, O., & Azap, S. (2026). Employing ChatGPT to improve high school students' writing skills by providing feedback on topic-specific writing tasks. *Computer Assisted Language Learning*, 1–23. https://doi.org/10.1080/09588221.2026.2613219

[16] Segaran, M., Engeness, I., & Gamlem, S. M. (2025). Lower secondary school teachers facilitation of students' writing process: A study of automated feedback and peer feedback approaches. *Computer Assisted Language Learning*, 1–26. https://doi.org/10.1080/09588221.2025.2539154

[17] Zhang, W. X., & Wu, Y. (2017). Assessment for learning: Constructing an automated essay evaluation system based on writing tasks. *Modern Educational Technology, 27*(3), 99–104.

[18] Lyu, M. (2015). Exploration and practice of intelligent assessment technology in the marking of large-scale English speaking examinations. *China Examinations*, (10), 51–57.

[19] Wang, Y. Y., Hu, W. T., & Guo, Q. (2026). GPT versus teacher feedback on English writing: An analysis of feedback characteristics and student engagement. *System, 138*, Article 103967.

[20] Song, Y., Xu, C.-L., & Mu, X.-X. (2024). Research on new classroom teaching evaluation and optimization enabled by generative artificial intelligence. *Modern Educational Technology, 34*(12), 27–36. https://doi.org/10.3969/j.issn.1009-8097.2024.12.003

[21] Wu, L. J. (2025). An empirical study of large language model-empowered English speaking instruction. *Modern Educational Technology*, (10), 33–41.

[22] Wen, Q. F., & Liang, M. C. (2024). Human-machine interactive negotiation competence: ChatGPT and foreign language education. *Foreign Language Teaching and Research, 56*(2), 286–296.

[23] Yang, L. R. (2024). Second language acquisition in the context of ChatGPT large language models. *Modern Foreign Languages, 47*(4), 578–585.

[24] Zheng, Y. Y. (2024). Applications of generative artificial intelligence in foreign language education: Key controversies and theoretical construction. *Foreign Language Education, 45*(6), 48–53.

[25] Xu, J. J., & Zhao, C. (2024). The roles of large language models in English language teaching. *Frontiers of Foreign Language Education Research, 7*(1), 3–10.

[26] Qin, H. W., & Lu, Y. F. (2024). Large language models and foreign language education: Application research based on language competence. *Foreign Language World*, (6), 37–44.

[27] Hong, H. Q., Qiao, Y. F., & Chai, Z. J. (2025). GenAI-empowered innovation in foreign language education: Internal logic and practical pathways. *Foreign Language World*, (3), 12–19.

[28] Xu, F. (2025). Constructing a teaching and research framework for AI-empowered cognitive development in English reading and writing. *Frontiers of Foreign Language Education Research, 8*(6), 38–48.

[29] Xu, J. J. (2025). Implementation strategies for personalized foreign language education in the context of large language models. *Foreign Language Teaching and Research, 57*(1), 81–91.

[30] Wang, X. M., & Zhou, M. J. (2024). Digital-intelligence literacy of university foreign language teachers: Connotation, framework, and development paths. *Foreign Language World*, (5), 33–40.

[31] Zhang, L. (2024). Exploring the application of large language models in foreign language teacher education and development research. *Foreign Language World*, (5), 25–32.

[32] Cui, Y., & Schunn, C. D. (2025). A survey of AI literacy among university teachers and students based on English academic writing tasks. *Foreign Languages and Their Teaching*, (6), 65–75.

[33] Tai, T. Y., & Chen, H. H. J. (2024). Improving elementary EFL speaking skills with generative AI chatbots: Exploring individual and paired interactions. *Computers & Education, 220*, Article 105112.

[34] Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning, 37*(1–2), 1–26. https://doi.org/10.1080/09588221.2021.2021241

[35] Yang, X. (2024). Enhancing teachers' self-efficacy beliefs in AI-based technology integration into English speaking teaching through a professional development program. *Teaching and Teacher Education, 145*, Article 104598.

[36] Li, S. S., Wu, X. M., Zhang, P., & Fan, Y. Z. (2025). Differences in effectiveness between human tutors and intelligent tutors: From the perspective of dynamic interactive affordances. *Open Education Research, 31*(2), 98–107. https://doi.org/10.13966/j.cnki.kfjyyj.2025.02.010

[37] Zheng, Z. Y., Fan, Q. Z., & Jia, W. (2024). The triple illusions of AI-empowered teacher development and the ways to overcome them. *China Educational Technology*, (7), 28–34.

[38] Fang, H. G., Kong, X. M., Liu, H. W., & Wang, X. C. (2024). Cooperative game and optimization strategies of human-machine collaborative education entities based on symbiosis theory. *e-Education Research, 45*(1), 21–27.

[39] Liu, X., Wen, Y., Yang, X., & Gao, Y. (2026). Generative AI-supported creative writing among primary school students: A study of creative writing performance, writing self-efficacy, and flow state. *Journal of Computer Assisted Learning, 42*(6), Article e70347. https://doi.org/10.1002/jcal.70347

[40] Muñoz Muñoz, B. C., Nassaji, H., & Bello Carrillo, F. (2025). ChatGPT-generated versus human direct corrective feedback on L2 writing. *System, 134*, Article 103805.

[41] Lu, Y., & Tang, X. Y. (2025). The hierarchical modes and progressive pathways of generative AI-empowered classroom teaching. *e-Education Research, 46*(6), 75–82.

[42] Zhong, M. (2025). Building an AI agent-based spoken English training model based on textbook characters. *Teaching and Management*, (11), 55–58.

[43] Xiao, L. (2023). Multiple biases in personalized learning in the intelligent era: A critical analysis of their risks. *Journal of Educational Studies, 19*(6), 55–66.

[44] Xing, T. Q., Di, P. C., & Chen, D. D. (2024). The constraints and transcendence of integrating large language models into primary school Chinese language teaching. *Teaching and Management*, (26), 33–38.

[45] Ministry of Education of the People's Republic of China. (2022). *Teacher digital literacy (JY/T 0646-2022)*.

[46] Xiao, W., Pian, Y., & Ma, X. M. (2025). The current status of AI application among primary and secondary school teachers and suggestions for training optimization: An empirical analysis based on large-scale survey data. *China Educational Technology*, (10), 114–122.

[47] Wang, Y. M., Wang, D., Liang, W. Y., & Liu, C. C. (2023). "Aladdin's lamp" or "Pandora's box": The potential and risks of ChatGPT in educational applications. *Modern Distance Education Research, 35*(2), 48–56.

[48] Wang, Y. M., Wang, X. Y., & Liu, C. C. (2024). Research on the ethical risk management framework of generative AI applications in education. *e-Education Research, 45*(10), 28–34.

[49] Wang, W., & Kou, Y. D. (2025). The potential crisis and countermeasures of popularizing AI education in primary and secondary schools: An analysis based on the triple structure of the digital divide. *Studies in Dialectics of Nature, 41*(11), 128–134.

[50] Lyu, X. Y., & Zou, T. L. (2026). The logical progression and implementation suggestions of AI ethics education. *China Educational Technology*, (3), 61–69.

[51] Xu, C. G. (2025). Human-AI collaborative writing engagement of foreign language learners based on generative artificial intelligence. *Foreign Languages and Their Teaching*, (2), 61–73.

[52] Ng, D. T. K. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology, 55*(5), 2067–2103.

Appendix A. Pilot protocol for the multi-tool, same-prompt observation

A.1 Material. The material is a Grade-5 second-semester English essay with nine planted errors (five tense, two spelling, and two collocation errors). The essay was teacher-composed for research purposes and contains no personal information of any minor.

A.2 Unified prompt

A.3 Procedure for a documented replication. The informal observations reported in Section 3.2 derive from single-pass use of the three tools by the author. A documented replication should proceed as follows. The same prompt is sent in fresh sessions to Doubao, iFLYTEK Spark, and ChatGPT, recording the model version and access date of each tool. Each tool is run three times. Two coders independently rate the outputs on the six dimensions in Table 1 and report inter-coder agreement, acknowledging that coders cannot be blinded to tool identity because of distinctive output formatting; disagreements are resolved through discussion. All raw outputs are archived for review.