Research on Strategies of AI-Empowered "PLANE" Writing Instruction in Senior High School English

Lili Zhang
The Middle School Attached to Ningxia University, Yinchuan, Ningxia, China

Abstract: Writing is the most heavily corrected and least effectively revised skill in senior high school English, because a teacher facing a hundred essays per task can point out errors but rarely has time to teach the rewriting. This paper develops a five-stage writing instruction framework, PLANE, in which artificial intelligence is embedded in a defined division of labour: Prompting, Learning, Assessing, Negotiating, and Editing. The framework is grounded in the cognitive process theory of writing and in recent empirical research on AI-generated feedback, which reports dependable gains in language accuracy alongside weak and variable gains in argumentation and discourse-level competence. Under the framework, the teacher designs prompts and rubrics before the task, students gather input from AI-curated model essays and corpora, an AI engine delivers first-round feedback on language form, the teacher and peers negotiate content and logic that the machine cannot judge, and students complete independent revision with version comparison. The framework was applied over twelve weeks in a Grade 11 class of 53 students in Yinchuan, Ningxia. Mean essay scores on the school rubric rose from 13.6 to 17.1 out of 25, grammar error density fell by roughly one third, and students recorded reasons for accepting or rejecting 58 percent of AI suggestions. Overreliance prevention, teacher workload, and boundaries of machine judgement are discussed.

Keywords: Artificial Intelligence; Writing Instruction; PLANE Model; Senior High School English; Feedback

1. Introduction

Of the four language skills tested in senior high school English, writing is the one where teacher effort and student gain diverge most. A teacher with two classes of around fifty students collects a hundred essays per writing task. Marking one essay properly, with marginal comments on logic, cohesion, and language, takes five to eight minutes, so the realistic load is ten hours per task. The teacher does it anyway, and two weeks later collects the next hundred essays whose errors look much the same. Students receive corrections, read the score, and file the paper away, because a corrected draft is not a revised draft.

Artificial intelligence changes the economics of this loop. Feedback engines built on large language models return sentence-level comments on grammar, word choice, and cohesion within seconds, can be prompted with a class rubric, and never tire of the hundredth essay. Whether such feedback actually improves writing is an empirical question rather than a commercial claim, and the research picture is two-sided: language accuracy improves fairly reliably, while argumentation and discourse-level gains stay weak unless a teacher adds human judgement on top [4, 6, 7]. This two-sided picture argues for a framework in which the machine does what it does well inside a structure the teacher controls.

This paper proposes such a framework, called PLANE after its five stages: Prompting, Learning, Assessing, Negotiating, and Editing. It was developed and applied in the author's school, the Middle School Attached to Ningxia University in Yinchuan. Section 2 reviews the theoretical basis and the empirical literature on AI writing feedback. Section 3 presents the PLANE framework and its rationale. Section 4 details the implementation strategies stage by stage. Section 5 reports a twelve-week classroom application, and Section 6 concludes with reflections and boundaries.

2. Theoretical Basis and Literature

2.1 Process Writing and Effective Strategies

The cognitive process theory of writing replaced the image of writing as transcription with the

image of planning, sentence generation, and revising as recursive mental activities [1]. For the classroom, the theory implies that instruction should intervene in the process, not only grade the product. Graham and Perin synthesised experimental research on adolescent writing and identified strategies with documented effects, among them prewriting activities, summarising, collaborative writing, specific product goals, sentence combining, and the study of models [2]. For second language writers, Hyland added that feedback works when it is specific, audience-aware, and distributed across drafts rather than concentrated on a final grade [3]. One thread in this tradition matters for the present design: several of the effective strategies, studying models, goal setting, and structured revision, are precisely the activities that AI tools can scale to a whole class.

2.2 AI Feedback for L2 Writing: Evidence and Limits

The empirical record on AI writing feedback has grown dense since 2023. Barrot analysed the potentials and pitfalls of ChatGPT for second language writing and concluded that the tool can support idea generation, text elaboration, and error correction, while carrying risks of overreliance and inaccurate corrections that demand teacher mediation [4]. Han and Li examined ChatGPT-supported teacher feedback in EFL classrooms and found that the combination, machine comments filtered and reframed by the teacher, produced better revision behaviour than either source alone [5]. Li followed 89 Chinese university students through twelve weeks of AI-assisted written corrective feedback and reported perceived gains across unity, support, cohesion, and language use with effect sizes between 0.77 and 1.57, alongside student reservations about the machine's limited rhetorical depth [6]. Al Ghaithi and Behforouz compared three feedback sources among 30 EFL learners, teacher feedback, EditGPT feedback, and general comments, and found the teacher-feedback group strongest, with the AI group still ahead of the control condition [7]. Mekheimer's randomised study with 60 postgraduate students found that a Grammarly-supported group significantly outperformed traditional instruction in writing proficiency and revised more often, with students asking for deeper comments beyond grammar [8]. A caution comes from a five-week study of 66 Turkish tenth-graders, where ChatGPT-integrated lessons produced less writing and vocabulary growth than conventional teaching, which the authors linked to unstructured tool use and learner confusion about how to respond to machine comments [9]. Read together, the studies support a division of labour: AI for speed, coverage, and language form; teacher and peers for argument, audience, and depth; and explicit structure for the interaction between the two.

3. The PLANE Framework

3.1 Definition of the Five Stages

PLANE names five stages of one writing cycle, and each stage fixes in advance who leads, the teacher, the students, or the machine. Prompting, the P, is preparation: the teacher converts the writing task into a set of designed prompts, one set for the students to elicit AI input and one embedded rubric for the feedback engine. Learning, the L, is input: students study AI-curated model essays, parallel texts, and corpus examples chosen at the class's measured level. Assessing, the A, is first-round feedback: the AI engine comments on the draft's language form against the embedded rubric, producing an error log. Negotiating, the N, is human feedback: the teacher holds short conferences and peers review in groups, focusing on argument, structure, and audience, and the class decides which machine comments stand and which are overruled. Editing, the E, is independent revision: each student rewrites, compares versions, and writes a short memo stating what was changed and why.

3.2 Rationale

The framework rests on the two-sided evidence in Section 2.2. The L and A stages occupy the territory where machine support has dependable effects, model study, coverage of every draft, and fast language correction [2, 6, 8]. The N stage guards the territory where machine judgement stays weak, argumentation, genre awareness, and rhetorical depth [6, 7], by routing those decisions through teacher conferencing and peer discussion, a combination that improved revision behaviour in classroom research [5]. The P stage exists because outcomes in the ten-grade study deteriorated when tool use lacked structure [9]; designing the prompts before the task is how structure is imposed. The E stage returns authority to the student, since revision carried out by the learner,

not correction delivered to the learner, is where writing develops [1, 3]. The acronym also carries its own image, a plane taking off, which students in the pilot class used to remember the order, and the metaphor is not entirely idle: each cycle is meant to gain altitude, from prompt to polished text.

4. Implementation Strategies

4.1 Prompting: Design Before the Task

One week before a writing task, the teacher prepares three artefacts: the task brief with audience and purpose stated in one sentence; a prompt library of six to eight prompts students may use in the L stage, such as asking a domestic chatbot like ERNIE Bot or DeepSeek for two model essays at different quality levels, or for ten topic-specific collocations with example sentences; and a rubric-prompt that instructs the feedback engine to score the draft against the school's five-dimension rubric and to mark error categories rather than rewrite sentences. Prompt literacy is taught once at the cycle's start: students learn to give the machine the genre, the length, the level, and the constraint, and to distrust fluent-sounding output that ignores the constraint [4].

4.2 Learning: Curated Input at Measured Level

In the L stage, students work through the prompt library for twenty minutes. Two model essays at different levels, rather than one, are always requested, and the class studies the gap between them with a comparison sheet: which thesis is arguable, which paragraph states its point first, which sentence pattern carries the argument. Corpus queries replace vocabulary lists; a student drafting an environmental essay asks for the ten most frequent verb-noun collocations in opinion essays on that topic and selects three to use. The teacher moves around the room, because a chatbot answer occasionally drifts off task and the teacher's quick judgement keeps the input on track [5, 9].

4.3 Assessing: Machine Feedback with an Error Log

Drafts are submitted to the feedback engine, which returns within one class period a marked copy and an error log organised by category: subject-verb agreement, articles, tense, word form, cohesion devices. The engine is configured to explain and locate, not to rewrite, so the student still performs the correction. Students transfer the log to a running sheet that accumulates across the semester, which turns individual feedback into a visible frequency distribution. The teacher reads the class's aggregated log, not each draft, and plans the next mini-lesson on the top three categories, a reversal of the marking routine that moved teacher attention from a hundred essays to one page of statistics.

4.4 Negotiating: Human Judgement on Content

The N stage runs as a fifteen-minute teacher conference per group of four plus one peer-review round. Conferences do not discuss grammar the machine has already covered; they ask three questions, what is your claim, who disagrees with it, and where does the reader lose the thread. In peer review, students exchange drafts with the machine comments attached and mark each comment as accept, reject, or doubt, giving a reason in one sentence. Rejected and doubted comments are discussed with the teacher, because a machine comment can be wrong, and learning to overrule it with reasons is itself feedback literacy [4, 5]. This stage implements the finding that rhetorical and discourse-level growth depends on human interaction the machine does not provide [6, 7].

4.5 Editing: Independent Revision and Version Memo

Each student revises alone, then submits both versions with a memo of at least four changes and their reasons. The platform's version comparison highlights deletions and insertions, and the memo forces an account of intent, not just a diff. Scores enter the portfolio together with the error-log trend, so the semester ends with each student's writing shown as a trajectory rather than a set of grades.

Two further routines support the E stage. The first is a five-minute silent reading of the final draft before submission, in which students read their own text aloud in a whisper and mark any line they stumble over; stumbled lines usually contain the cohesion and agreement errors the eye tolerates and the ear catches. The second is a monthly excerpt wall. Four memos are selected each month, two from strong writers and two from students whose error logs fell fastest, and posted anonymously with the teacher's one-line

comment on what changed between versions. The wall costs nothing to run and gives the class a shared sense of what revision looks like when it succeeds, which matters for students who have never seen a real second draft, only corrected first drafts. Table 1 summarises the cycle.

Table 1. The PLANE Cycle: Stages, AI Functions, Teacher Roles, and Student Outputs

Stage	AI Function	Teacher Role	Student Output
Prompting	Holds prompt library and embedded rubric	Design task brief, prompts, and rubric	Understood task with audience and purpose
Learning	Curates model essays and corpus input	Monitor input quality, guide comparison	Notes on model gap, selected collocations
Assessing	Marks drafts by error category, builds logs	Read aggregated log, plan mini-lesson	Corrected draft, updated error sheet
Negotiating	Provides comments for accept-reject judgement	Conference on claim, objection, and flow	Marked comments with one-line reasons
Editing	Version comparison display	Score revision and memo	Final draft with revision memo

5. A Twelve-Week Classroom Application

5.1 Setting and Procedure

The framework was applied in the author's school during the autumn semester of 2025 in Class 5 of Grade 11, with 53 students, over one argumentative writing unit of twelve weeks. One full PLANE cycle ran every two weeks, giving six cycles: prompt release and the L stage in the first period, drafting and machine assessment in the second, negotiation and editing in the third, with the memo due before the next prompt. The school writing rubric scores content, organisation, language, and mechanics out of 25 points, and the same rubric was embedded in the feedback engine's prompt so that machine comments and human scores spoke the same language. In week 1 and week 12, all students wrote a 120-word opinion essay on parallel topics under examination conditions, scored blind by two teachers; the two raters' scores differed by more than two points on four essays, which were re-scored after discussion. Across the cycles, records were kept of essay scores, error-log categories, and the accept-reject decisions on machine comments.

5.2 Results and Observations

The mean rubric score rose from 13.6 in week 1 to 17.1 in week 12. The gain concentrated where the design predicted it: language rose from 4.1 to 6.0 out of 8, mechanics from 2.9 to 4.1 out of 5, and organisation from 3.4 to 4.4 out of 7, while content, the dimension the machine does not assess, moved more modestly from 3.2 to 3.9 out of 5. Error-log totals fell from an average of 11.4 language errors per essay in cycle one to 7.6 in cycle six, a reduction of about one third, and the three persistent categories across the class were articles, subject-verb agreement, and cohesion devices, in that order. Students recorded reasons for 58 percent of machine suggestions; of these, 71 percent were accepted, 22 percent rejected with a reason, and 7 percent marked as doubt and raised in conference. The rejection rate is worth attention, since it shows the negotiation stage doing real work rather than rubber-stamping the machine [5, 6]. In the questionnaire at week 12, 44 of 53 students agreed that the version memo helped them see their own improvement, 9 were neutral, and none disagreed. Eleven students wrote in the open response that the machine sometimes corrected a sentence they had intended as emphasis, and asked for clearer rules on when a native-like rewrite should be refused; this echoes the reported mismatch between machine corrections and rhetorical intent [4, 7]. The teacher's marking time per task fell from about ten hours to about three, spent on conferencing and reading memos rather than correcting language.

One cycle illustrates how the pieces interact. In cycle four the prompt asked students to argue whether senior high schools should shorten the winter holiday. In the L stage, most pairs requested model essays at two levels and the comparison sheets filled quickly; one group's chatbot produced a model essay of 400 words despite the length constraint, which the class discussed as a live example of why the constraint is stated in the prompt. In the A stage, the aggregated error log showed cohesion devices at the top for the first time, overtaking articles, so the next mini-lesson shifted to

linking words. In the N stage, one conference centred on an essay whose three body paragraphs all argued the same point in different words; the machine had marked it as well organised, because its cohesion comments track connectives, not logic, and the peer group caught the repetition within four minutes. The student's memo recorded the fix, one paragraph rewritten as a counter-argument, and the final version scored 20 of 25, the highest of that cycle. The episode captures the division of labour the framework intends: the machine kept every sentence accountable for form, and the humans kept the argument honest.

The usual cautions apply. One class, one unit, no control group, and a school rubric limit the claims; the week-12 topic may have been marginally easier despite the parallel design; and rater discussion on four essays, though documented, weakens strict blindness. The study also says nothing about retention after the framework is withdrawn, a question left to the next semester.

6. Conclusion and Reflections

This paper proposed PLANE, a five-stage framework that places AI inside senior high school writing instruction under a fixed division of labour, and reported a twelve-week application in which essay scores, error reduction, and student engagement all moved in the intended direction while teacher marking time fell. Three reflections temper the optimism. First, the machine earned its place in the L and A stages by doing what evidence says it does well, covering every draft with fast language feedback, and was kept out of the N stage where evidence says its judgement is weak; the framework's value lies in this routing, not in any single tool [4, 6, 7]. Second, structure preceded the tool. Prompt design, the rubric-prompt, the comparison sheet, and the memo existed on paper before any chatbot was opened, because unstructured tool use has measurably underperformed conventional teaching [9]. Third, overreliance is a live risk that the framework manages but does not eliminate; the accept-reject reason record made students argue with the machine weekly, yet eleven still asked for clearer refusal rules, which is where teacher judgement stays irreplaceable [5]. For schools considering adoption, a realistic starting point is one class, one unit, six cycles, with the error-log aggregate reviewed at each cycle's end. The machine did not teach the writing; it made the teaching visible in time to matter.

References

[1] Flower L, Hayes J R. A cognitive process theory of writing. College Composition and Communication, 1981, 32(4): 365-387.

[2] Graham S, Perin D. Writing Next: Effective Strategies to Improve Writing of Adolescents in Middle and High Schools. Washington, DC: Alliance for Excellent Education, 2007.

[3] Hyland K. Second Language Writing. Cambridge: Cambridge University Press, 2003.

[4] Barrot J S. Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 2023, 57: 100745.

[5] Han J, Li M. Exploring ChatGPT-supported teacher feedback in the EFL context. System, 2024, 126: 103502.

[6] Li S. Perceptions and effectiveness of AI-assisted written corrective feedback: A case study of Chinese EFL university students. Journal of Education and Training Studies, 2025, 13(4).

[7] Al Ghaithi A, Behforouz B. Effects of interaction with AI-assisted writing evaluation on EFL students' writing performance. Knowledge Management & E-Learning, 2025, 17(2): 206-224.

[8] Mekheimer M. Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. Discover Education, 2025, 4: 170.

[9] Sapan M, Uzun L. The effect of ChatGPT-integrated English teaching on high school EFL learners' writing skills and vocabulary development. International Journal of Education in Mathematics, Science and Technology, 2024, 12(6): 1657-1677.

