Apollo Journals
Social Sciences and Humanities Inquiry
SSHI
Vol. 01 Issue 02 2026

Article

Basic Principles and Practical Paths of LLM-Empowered Senior High School English Writing Evaluation in China: A Conceptual Framework

Jiang Li [a,b]

[a] *College of Education, Zhejiang Normal University, Jinhua 321004, China; Email: 1292451789@qq.com*

[b] *Zhengzhou No. 36 Middle School, Zhengzhou 450042, China*

Abstract: This paper adopts a conceptual analysis approach based on policy document interpretation and narrative literature synthesis to explore large language model empowered senior high school English writing evaluation within the Chinese educational context. Driven by national digital transformation of education and the 2020 Overall Plan for Deepening the Reform of Educational Evaluation in the New Era, generative large language models deliver innovative technical support to resolve prominent drawbacks of traditional writing assessment including superficial evaluation dimensions, single evaluation subjects, rigid marking mechanisms and delayed interactive feedback. Different from conventional automated writing evaluation tools limited to surface linguistic scoring, large language models generate conversational feedback, support adjustable evaluation standards through prompt design and conduct comprehensive discourse assessment covering language competence, cultural awareness, thinking quality and learning ability which forms the core innovation of this research. This study systematically elaborates two guiding orientations for human machine collaborative evaluation including disciplinary core literacy orientation and digital intelligent diversified evaluation orientation and constructs three targeted practical paths matching real senior high school teaching scenarios. The three paths cover teacher led evaluation guided by curriculum standards, visualized learner profiling for targeted instruction and cross time and space personalized writing guidance. Supplementary supporting measures regarding academic integrity, minor data compliance, balanced digital resource allocation and teacher artificial intelligence literacy cultivation are further supplemented to improve practical operability. The established framework helps shift assessment logic from score oriented judgment to literacy oriented developmental guidance. Restrictions such as algorithm bias, student over dependence on artificial intelligence and data ethical risks hinder stable large scale application of model based evaluation and relevant discussions are fully presented in the limitations section. This research provides theoretical frameworks and operable classroom strategies for intelligent reform of senior high school English writing evaluation in China.

Keywords: Large Language Models; Automated Writing Evaluation; Formative Assessment; Senior High School EFL Writing; Disciplinary Core Literacy

Received: 3 August 2026
Revised: 21 August 2026
Accepted: 10 September 2026
Published: 13 October 2026

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/).

1. Research Background

In October 2020, the Central Committee of the Communist Party of China and the State Council officially issued the *Overall Plan for Deepening the Reform of Educational Evaluation in the New Era* (Central Committee of CPC & State Council, 2020),which clearly emphasizes the importance of innovating educational evaluation tools and methods, actively applying modern information technologies such as artificial intelligence, big data and intelligent algorithms, and building a full-cycle, multi-dimensional and developmental student evaluation system. The policy points out that educational evaluation should break the single score-oriented evaluation mode, pay attention to students' whole-process growth and personalized development, and realize the organic integration of teaching, learning and evaluation, which provides clear policy guidance for the intelligent reform of basic English education evaluation. In the context of comprehensive educational digitalization, generative artificial intelligence technology represented by ChatGPT, DeepSeek and other large language models has achieved iterative breakthroughs in natural language processing, text

understanding, logical reasoning and content generation, bringing disruptive innovation to the traditional teaching evaluation system (Lu et al., 2023).

English writing is a high-level cognitive activity that integrates language accumulation, logical organization, cultural understanding and creative expression, and it is also a key difficulty and core teaching link in senior high school English teaching. Different from listening, speaking and reading skills that focus on information input and reception, writing is a comprehensive output process that requires students to integrate language knowledge, sort out logical relationships and convey emotional attitudes and cultural views. Nevertheless, traditional senior high school English writing evaluation has long relied entirely on teachers' manual marking and empirical judgment. Under the teaching situation of large class size and heavy teaching tasks, teachers can only complete superficial scoring and simple error marking, and it is difficult to provide systematic, targeted and developmental personalized guidance for each student's writing problems (Li & Zhu, 2024). In contrast, large language models have powerful natural language analysis, multi-dimensional evaluation and instant feedback capabilities, which can efficiently complete refined evaluation of students' compositions, dig deep into students' writing advantages and existing deficiencies, and generate targeted revision suggestions and learning guidance. Based on the policy background of educational evaluation reform and the practical dilemma of traditional writing evaluation, this paper takes LLM-empowered senior high school English writing evaluation as the research object, systematically explores its basic evaluation principles and feasible practical paths, so as to provide effective practical references for promoting the intelligent, refined and developmental transformation of senior high school English writing evaluation.

1.1. Dilemmas of Traditional Senior High School English Writing Evaluation

Traditional teacher-led manual writing evaluation has long been the mainstream mode of senior high school English writing teaching evaluation, which can ensure the basic fairness and standardization of evaluation results under the condition of limited educational resources, and effectively supervise students' mastery of basic English language norms such as vocabulary, grammar and sentence patterns. However, under the guidance of the core literacy-oriented *General High School English Curriculum Standards (2017 Edition, Revised 2020)*, the traditional writing evaluation mode has been difficult to adapt to the new requirements of high-quality English teaching, and there are three prominent practical dilemmas in actual teaching operation (Ministry of Education of the People's Republic of China, 2020).

First of all, the evaluation content is superficial and single, lacking multi-dimensional literacy assessment. Traditional writing evaluation overly focuses on superficial linguistic error inspection, mainly scoring and correcting students' vocabulary spelling, grammatical tense, sentence pattern collocation and other basic errors, while ignoring the in-depth evaluation of core literacy indicators such as discourse coherence, textual logic, structural layout, thematic expression, cultural connotation and innovative thinking. Restricted by limited class hours, heavy marking tasks and empirical evaluation thinking, most teachers cannot conduct comprehensive and detailed analysis and evaluation on the overall quality of students' compositions. This single error-oriented evaluation method leads to the formalization and simplification of writing evaluation, cannot effectively feed back students' real writing level and literacy development status, and also makes students form the wrong writing cognition of "only paying attention to no grammatical errors", which seriously restricts the improvement of students' comprehensive writing literacy.

Secondly, the evaluation subject is single and closed, lacking student independent participation. For a long time, senior high school English writing evaluation has formed a single evaluation pattern dominated entirely by teachers. Teachers are not only the implementers of writing teaching, but also the sole evaluators of writing results, while students are always in a passive accepted evaluation state, lacking independent participation, self-reflection and mutual evaluation experience. Although some front-line teachers try to introduce student self-evaluation and peer mutual evaluation into writing teaching practice, due to students' insufficient English language literacy, lack of systematic evaluation criteria and weak discourse analysis ability, the implementation effect of student participation evaluation is poor, and it is difficult to form effective evaluation feedback (Zhang & Yang, 2023). The long-term absence of student main body participation makes writing evaluation lose the function of promoting student self-reflection and autonomous improvement, and greatly weakens the developmental educational value of writing evaluation.

Finally, the evaluation method is rigid and inefficient, lacking real-time interactive feedback. Affected by the traditional teaching process arrangement, most senior high school English writing evaluations adopt the centralized after-class marking mode. Teachers uniformly correct students' compositions after class, and return the compositions with simple error marks and scores after a long correction cycle. Most evaluation feedback belongs to single indirect feedback, which only marks wrong sentences and errors without providing targeted correction methods, logical analysis and detailed explanation of error causes (Bitchener, 2012). This restricted feedback type corresponds to what Ellis (2009) classified as indirect corrective feedback, which provides no explicit correction and thus places a heavy interpretive burden on learners; by contrast, direct and metalinguistic feedback types are more likely to promote noticing and subsequent revision. The long feedback cycle and single feedback form make it difficult for students to remember their writing ideas and existing problems at that time, resulting in low efficiency of feedback absorption. It is impossible

https://doi.org/10.67728/sshi.2026.010

to form an effective closed-loop interaction of "writing-evaluation-revision-improvement", which hinders the continuous iterative improvement of students' writing ability.

1.2. Solutions to Evaluation Dilemmas Empowered by Digital Intelligence Technology

The rapid iteration and popularization of artificial intelligence and digital intelligence education technology provides a new breakthrough for solving the above practical dilemmas of traditional English writing evaluation. In particular, the continuous maturity of large language model technology based on generative pre-training, human feedback reinforcement learning and chain-of-thought reasoning has achieved qualitative leaps in machine natural language understanding, text analysis, logical judgment and personalized content generation, providing solid and reliable technical support for the intelligent transformation of English writing evaluation (Lu et al., 2023). Different from traditional single intelligent scoring software, modern large language models have strong contextual understanding and independent reasoning ability, can adapt to different writing genres, writing themes and student literacy levels, and realize refined and differentiated evaluation of English compositions.

Specifically, the generative pre-training technology of LLMs enables the model to master massive English corpus and writing evaluation rules through ultra-large-scale text training, efficiently process long-sequence English composition texts, and generate multi-dimensional and hierarchical evaluation feedback including text comments, error analysis, revision suggestions and optimization directions according to customized evaluation criteria, which greatly enriches the dimension and depth of traditional writing evaluation. The human feedback reinforcement learning technology further optimizes the model output effect. Through long-term human expert instruction training and proximal policy optimization algorithms, LLMs can continuously calibrate evaluation errors, adapt to the evaluation standards of senior high school English writing, and serve as an intelligent third-party evaluation agent to provide adaptive and developmental evaluation feedback for students' writing (Zheng et al., 2023). In addition, the chain-of-thought reasoning technology empowers LLMs to disassemble complex writing evaluation tasks step by step, provide referential evaluation superficial linguistic errors and deep logical and structural problems in compositions, generate standardized and logical evaluation results, and carry out interactive meta-linguistic feedback, guiding students to deeply explore the internal causes of writing errors, so as to improve students' writing cognition and autonomous revision ability fundamentally (Lu et al., 2023).

Conventional automated writing evaluation tools operate with fixed preset scoring algorithms that only detect superficial lexical and grammatical errors. Such tools cannot produce open-ended generative feedback, support interactive follow-up questioning, adjust evaluation standards through prompt design, or conduct multi-layered assessment covering discourse logic and cross-cultural expression (Liu et al., 2022). In contrast, large language models deliver context-aware conversational revision suggestions and flexibly adjust evaluation dimensions to match the disciplinary core literacy requirements stated in curriculum documents. Empirical research confirms that large language models achieve stronger performance in discourse-level writing assessment and personalized feedback delivery compared with earlier automated scoring tools (Escalante et al., 2023). Domestic research on artificial intelligence assisted foreign language teaching further indicates that the dynamic generative capacity of large language models fills the long-standing functional gaps of traditional automated writing evaluation tools in Chinese English teaching contexts (Cai, 2025; Zhang & Hong, 2023). This fundamental functional gap between two generations of evaluation technology forms the core innovative contribution of the present research.

2. Basic Principles of Senior High School English Writing Evaluation Empowered by Large Language Models

2.1. Aligning with Disciplinary Core Literacy to Accurately Orient Evaluation Contents

English curriculum teaching and evaluation in the new era must always adhere to the student-oriented educational concept, take the cultivation of students' disciplinary core literacy as the fundamental goal of all teaching and evaluation activities, and realize the transformation of evaluation from "score-oriented judgment" to "literacy-oriented development" (Chen & Hu, 2022). As an important part of English teaching evaluation, writing evaluation must also take core literacy cultivation as the core orientation. LLM-empowered intelligent writing evaluation breaks the limitation of traditional single language evaluation, and builds a full-dimensional evaluation system covering language competence, cultural awareness, thinking quality and learning ability, realizing comprehensive evaluation and developmental guidance of students' writing comprehensive literacy.

In terms of language competence cultivation, language competence is the basic foundation of English writing and the primary evaluation dimension of writing teaching. Traditional manual evaluation only focuses on the standardization of students' basic language application, while LLMs rely on massive professional English corpus and powerful language analysis ability to conduct all-round and in-depth evaluation of students' writing. It can not only provide referential evaluation basic errors such as vocabulary misuse and grammatical confusion, but also evaluate the richness of vocabulary application, the diversity of sentence pattern selection, the fluency of discourse expression and the rationality of stylistic specification (Yang, 2024). The human-machine collaborative

https://doi.org/10.67728/sshi.2026.010

evaluation mode expands the evaluation scope from superficial language knowledge inspection to in-depth language application ability evaluation, effectively promotes the integrated improvement of students' writing language accuracy, fluency and appropriateness, and lays a solid language foundation for high-quality English writing output.

Relevant empirical research confirms that large language models deliver multi-layered linguistic feedback covering lexical richness syntactic diversity and discourse coherence which traditional manual marking cannot fully realize (Guo et al., 2023). Learners can receive targeted revision advice on word selection sentence transformation and cohesive logic through model evaluation and form clear cognition of personal linguistic weaknesses in writing practice (Xu et al., 2024). Such fine-grained linguistic assessment makes up for the limited marking energy of frontline teachers under large class teaching conditions and realizes continuous optimization of students' comprehensive language application competence in long-term writing training.

In terms of thinking quality cultivation, writing is an important carrier of cultivating students' logical thinking, critical thinking and innovative thinking. Traditional writing evaluation rarely involves the evaluation of students' thinking level, resulting in the disconnection between writing teaching and thinking cultivation. LLM intelligent evaluation can build a systematic thinking training scaffold for students' writing growth (Gu, 2018). In the evaluation process, teachers can combine LLM's logical analysis function to guide students to sort out writing framework and thematic context, carry out critical questioning on unreasonable content and confused logic in writing, and encourage students to carry out innovative exploration in material selection, viewpoint expression and content layout. Through continuous evaluation and guidance, it helps students form rigorous logical thinking habits, develop independent critical thinking ability and innovative expressive thinking, and realize the synchronous improvement of writing level and thinking quality.

Reasonable human-machine interaction during writing evaluation can drive students to conduct repeated questioning and logical deduction on text content which becomes an effective carrier for critical thinking cultivation (Cai, 2025). Classroom practice supported by intelligent tools transfers repetitive knowledge checking tasks to models and leaves sufficient classroom time for teachers to organize high-order thinking discussion and text critical analysis (Kong, 2024). This differentiated task allocation reverses the tendency of superficial skill training in traditional writing teaching and highlights the core position of logical reasoning and innovative expression in writing literacy cultivation.

In terms of cultural awareness cultivation, English writing is an important way for students to output cross-cultural cognition and express cultural views. Traditional writing evaluation ignores the evaluation of students' cultural cognition and cross-cultural expression ability, resulting in students' writing lacking cultural connotation and humanistic value. LLMs draw on broad multilingual and multicultural training corpora accumulated in pre-training stages and rich cross-cultural case resources, which can provide referential evaluation the appropriateness of students' application of Western cultural knowledge and the rationality of cross-cultural viewpoint expression in writing, and judge students' inclusive and dialectical understanding of Chinese and foreign diverse cultures. Through intelligent evaluation and feedback, LLMs help students correct cross-cultural expression errors, establish correct cross-cultural values, enhance cultural confidence and cross-cultural communication awareness, and effectively promote the internalization of students' English cultural literacy

Large language models are limited by unbalanced training corpus distribution and tend to show bias toward western cultural narratives during evaluation which requires teachers to undertake the final judgment of cultural expression rationality in human-machine collaborative assessment (Liu et al., 2025). The evaluation link can set specific scoring dimensions for Chinese cultural narrative expression in student compositions and guide learners to deliver inclusive and dialectic cross-cultural viewpoints through targeted model feedback. Such design balances the information advantages of artificial intelligence and the value leading function of teachers in cultural literacy education.

In terms of learning ability cultivation, traditional writing evaluation is dominated by teachers' one-way evaluation, and students can only passively accept evaluation results, lacking independent reflection and autonomous learning space. The application of LLM intelligent evaluation effectively breaks this dilemma. LLMs provide students with diversified learning resources, instant evaluation feedback and targeted revision suggestions, enabling students to independently complete writing self-inspection, error analysis and content optimization. In the long-term independent evaluation and revision process, students can continuously summarize writing experience, optimize writing strategies, form good autonomous learning and reflective learning habits, and effectively improve their independent learning ability and lifelong English literacy development potential.

Learners form stable positive willingness to use intelligent evaluation tools when they perceive obvious promotion of personal writing efficiency which conforms to the core logic of technology acceptance model in foreign language learning research (Wang & Wang, 2024). Continuous independent revision supported by model feedback promotes multi-dimensional learning engagement including behavioral cognitive and emotional participation and deepens the formation of self-regulated writing strategies (Xu, 2025). The whole process weakens students' passive learning state under single teacher evaluation and constructs a circulating self-improvement mechanism for long-term foreign language writing learning.

https://doi.org/10.67728/sshi.2026.010

2.2. Leveraging Digital Intelligence Advantages to Implement Diversified Evaluation Methods

Digital intelligence technology has completely changed the single and solidified pattern of traditional teaching evaluation, constructed a new human-machine collaborative multi-subject evaluation paradigm, and realized the innovative transformation of English writing evaluation from single teacher empirical evaluation to multi-subject negotiated evidence-based evaluation (Liu & Chen, 2024). LLM intelligent empowerment effectively makes up for the deficiencies of traditional evaluation methods, and promotes the innovative development of English writing evaluation modes from four core dimensions, realizing the comprehensive upgrading of evaluation timeliness, objectivity, comprehensiveness and individuation.

Firstly, it promotes the transformation from result-oriented static evaluation to process-oriented dynamic evaluation. Traditional writing evaluation belongs to typical result-oriented evaluation, which only focuses on the final writing score and composition quality, ignores students' writing process, phased progress and growth changes, and cannot capture students' problems and deficiencies in the writing process. LLM-empowered intelligent evaluation realizes real-time human-machine dialogue and whole-process phased evaluation. Students can initiate intelligent evaluation at any stage of writing conception, draft writing and revision, obtain real-time feedback guidance, and dynamically adjust writing ideas and content according to evaluation suggestions. Teachers can also grasp students' phased learning progress, growth advantages and existing problems in real time through LLM process data, realize precise teaching guidance in the zone of proximal development, and promote students' progressive growth in writing ability.

Secondly, LLMs can reduce human scoring bias caused by teacher fatigue, personal preference and inconsistent marking standards, yet they introduce new types of variability arising from prompt sensitivity, random model output and training data bias. Instead of realizing fully objective evaluation, LLMs provide standardized quantitative evidence to assist teachers' comprehensive, balanced judgment.Traditional teacher evaluation is easily affected by personal teaching experience, subjective preference and emotional state, resulting in subjective deviation and inconsistent evaluation standards, which affects the fairness and scientificity of evaluation results. With powerful big data analysis and precise quantitative measurement capabilities, LLMs can carry out quantitative and qualitative comprehensive evaluation of students' writing, generate objective and accurate evaluation data and analysis reports, and provide reliable data evidence for student self-evaluation, peer mutual evaluation and teacher comprehensive evaluation (Zhou, 2017). Evidence-based intelligent evaluation effectively avoids subjective empirical errors, accurately identifies students' writing growth points and ability deficiencies, and may enhance the rigor of evaluation under standardized use, rigor and effectiveness of writing evaluation.

Thirdly, it breaks through temporal and spatial constraints and builds an integrated teaching-evaluation dynamic system. Traditional writing evaluation is limited by classroom teaching time and offline teaching space, and can only be carried out in centralized classroom teaching and after-class marking links, lacking continuous and long-term evaluation support. LLM intelligent evaluation can realize anytime and anywhere online evaluation and interactive revision, breaking the temporal and spatial limitations of traditional evaluation. Students can complete writing evaluation and autonomous revision after class, on holidays and in daily independent learning scenarios, realizing real-time integration of teaching, learning and evaluation, forming a continuous closed-loop of learning-evaluation-improvement, and providing long-term and stable dynamic support for students' sustained writing ability improvement.

Fourthly, it effectively stimulates students' writing motivation and autonomous learning awareness. Traditional single scoring and error-marking evaluation mode is boring and passive, which is easy to make students resist writing learning and lack writing enthusiasm. LLMs can provide instant, targeted and positive diversified feedback, including detailed advantage comments, personalized problem analysis and operable revision suggestions. The diversified and encouraging evaluation feedback can effectively activate students' writing interest and subjective initiative, change students' passive writing attitude, help students establish writing self-confidence, cultivate good habits of independent writing, self-inspection and self-optimization, and realize the continuous and progressive improvement of comprehensive writing ability (Zhang et al., 2024).

The above analysis sorts out the three prominent practical dilemmas of traditional English writing evaluation, two core guiding orientations of large language model empowered assessment, and three targeted classroom implementation paths. To clarify the internal logical correspondence among dilemma diagnosis guiding principles and practical operations intuitively, a unified summary table is presented as follows (Table 1).

https://doi.org/10.67728/sshi.2026.010

Table 1. Dilemmas-Principles-Paths Correspondence Summary Table. Source: Author.

Traditional Evaluation Dilemmas	Matching Guiding Principles	Corresponding Practical Paths	Specific Teacher Classroom Operations
1. Superficial, single evaluation dimensions; overemphasis on grammatical errors while ignoring multi-dimensional disciplinary core literacy assessment	Disciplinary core literacy orientation	Path 1: Adhere to curriculum standards and leverage teachers' professional judgment	Co-construct multi-dimensional literacy evaluation rubrics with students; calibrate biased LLM cultural & logical assessment outputs; conduct targeted literacy training based on AI error statistics
2. Single, closed evaluator subject; students passively accept scores without self-assessment and peer feedback mechanisms	Digital-intelligent diversified evaluation orientation	Path 2: Build visualized learner profiling to realize precise differentiated teaching	Extract quantitative writing data via LLMs; generate traceable student writing archives; design stratified writing tasks matching individual learner strengths and deficiencies
3. Rigid offline marking; long feedback cycles lacking real-time interactive revision; failure to form writing-evaluation-revision closed loop	Digital-intelligent diversified evaluation orientation	Path 3: Break temporal-spatial limits to deliver whole-process personalized writing instruction	Establish after-class AI self-evaluation channels; design standardized human-AI collaborative writing workflow; track long-term writing growth trends through continuous LLM data recording
Supplementary hidden risks: AI academic integrity risks, minor student data security, uneven digital resource access, insufficient teacher AI literacy	Dual-principle collaborative constraint mechanism	Supplementary guarantee measures (Section 3.4)	Formulate classroom AI writing supervision rules; adopt compliant school-localized AI platforms; organize regular teacher AI evaluation training; adjust implementation schemes for resource-limited schools

3. Practical Paths of Senior High School English Writing Evaluation Empowered by Large Language Models

3.1. Focusing on Curriculum Standards and Leveraging Teachers' Distinctive Professional Judgment in Human-Machine Collaboration

Standardized and scientific evaluation criteria are the basic premise to ensure the efficient and high-quality implementation of LLM-empowered English writing evaluation. Although large language models have super powerful knowledge reserve, language analysis and data computing capabilities, their evaluation results belong to probabilistic fuzzy output based on algorithm training, which has inherent limitations such as insufficient targeted logical pertinence and lack of humanistic emotional judgment and creative thinking identification (Sun, 2021). Intelligent tools can only serve as auxiliary evaluation carriers, and cannot replace teachers' subjective judgment, humanistic guidance and value leadership in teaching evaluation. It is necessary to clarify operable implementation steps to translate human-machine collaborative evaluation theories into daily classroom practice. A complete human-AI collaborative writing evaluation workflow is designed as follows. First, teachers formulate multi-dimensional evaluation rubrics aligned with curriculum standards. Second, students finish drafts independently in class. Third, large language models conduct initial automatic scoring, error analysis and draft feedback. Fourth, teachers review outputs generated by large language models, adjust biased evaluation results and supplement humanistic value judgment. Fifth, students revise compositions according to combined feedback from artificial intelligence and teachers. Sixth, large language models carry out secondary assessment for revised versions to track student writing improvement.

To avoid confusion over subject division in actual operation, clear responsibility boundaries between teachers and intelligent tools should also be defined. Teacher responsibilities cover setting unified evaluation standards, judging cultural connotation and creative value presented in student writing, and guiding students to form rigorous logical thinking. Auxiliary work undertaken by large language models includes surface linguistic error detection, quantitative statistical analysis of writing data, and provision of preliminary targeted revision suggestions for learners.

Specifically, teachers need to deeply interpret the core literacy objectives and writing teaching requirements of the senior high school English curriculum standards, combine the cognitive characteristics, learning foundation and practical writing level of senior high school students, clarify hierarchical and phased writing teaching objectives, and construct a systematic, standardized and operable multi-dimensional writing evaluation framework covering genre standardization, vocabulary application ability, grammatical structure diversity, textual logic coherence, thematic expression and cultural connotation. On this basis, teachers can negotiate and

https://doi.org/10.67728/sshi.2026.010

optimize detailed evaluation rules with students according to actual teaching conditions, ensure that the evaluation criteria conform to teaching reality and student characteristics, and use LLMs to provide adaptive intelligent teaching and evaluation support. Meanwhile, teachers can rely on LLM big data statistical function, summarize students' common writing errors and typical problems in stages, formulate targeted teaching optimization strategies and special writing training plans, accurately make up for students' writing deficiencies, and promote the overall and balanced improvement of students' English writing level.

3.2. Empowering Accurate Learner Profilings for Progressive Learning

Precision teaching and personalized education are the core development directions of current basic education reform, and student learner profiling technology is an important technical means to realize evidence-based teaching evaluation and precise teaching (Hu & Lin, 2019). learner profiling technology relies on multi-modal learning data collection, quantitative analysis and label visualization, to systematically sort out students' learning characteristics, ability advantages and existing deficiencies, providing accurate data support for differentiated teaching. In LLM-empowered senior high school English writing evaluation, large language models can realize full-dimensional in-depth analysis of students' writing works, comprehensively collect and sort out quantitative data such as students' linguistic expression accuracy, vocabulary richness, sentence pattern diversity, textual structure rationality and logical thinking clarity, and form systematic, visualized and personalized student writing learner profilings.

Teachers can carry out accurate label classification and feature summary according to LLM's writing analysis results, form quantifiable, storable and traceable student writing learning archives, accurately grasp each student's personalized writing advantages and typical deficiencies, and dynamically adjust writing teaching strategies and training focus according to portrait data. For common writing problems such as single sentence pattern, insufficient scene description, vague thematic expression and confused logical structure existing in different students, teachers can design hierarchical, targeted and differentiated writing training tasks, realize precise matching of students' learning situation and teaching tasks, reshape the single exam-oriented writing teaching mode to a certain extent, and provide one-to-one personalized growth guidance for every student's writing ability improvement.

3.3. Breaking Temporal and Spatial Limitations to Realize Personalized Writing Teaching

The integration of large language model technology and English writing teaching breaks the inertial limitations of traditional teacher-dominated, classroom-restricted writing evaluation mode, builds a new collaborative interactive ecosystem of teachers, students and intelligent tools, and realizes personalized writing teaching and whole-process developmental evaluation across time and space. In the intelligent teaching environment, students can break the restrictions of classroom teaching time and offline space, carry out real-time writing evaluation, error checking and content revision through LLMs in after-class learning, holiday review and daily autonomous learning scenarios, obtain customized evaluation feedback, targeted revision suggestions and expanded learning resources, and greatly improve the autonomy, pertinence and scientificity of writing self-evaluation and autonomous learning.

In the process of human-machine collaborative evaluation and revision, students can give full play to their subjective cognitive advantages, rationally judge the rationality and applicability of LLM intelligent feedback results, independently screen effective revision suggestions, and complete writing optimization and content innovation. This process can effectively stimulate students' writing self-efficacy and autonomous learning motivation, and promote their cognitive ability and emotional quality growth. At the same time, teachers can integrate students' long-term personalized writing evaluation data, build continuous process-based writing learning archives, accurately analyze the changing rules, growth trends and internal restrictive factors of students' writing ability in the whole learning stage, design gradient progressive writing tasks according to individual student differences, fully tap students' writing potential, and realize the long-term and systematic improvement of students' English writing comprehensive literacy.

3.4. Supplementary Practical Guarantees

The three practical paths outlined above clarify core human machine collaborative evaluation operations for daily writing teaching. Stable long term implementation of such intelligent evaluation modes still relies on supporting institutional and operational safeguards covering academic integrity data compliance and balanced resource allocation. This section sorts out three essential guarantee dimensions to make the whole evaluation framework more complete and operable.

3.4.1. Academic Integrity and Writing Construct Validity Control

Widespread access to large language models creates new hidden risks to the authenticity of student writing work. Learners may rely on artificial intelligence to finish full compositions or rewrite most text content which weakens the ability of writing evaluation to reflect real student literacy level. Targeted classroom supervision mechanisms need to be designed to resolve such risks. Schools can implement in class timed writing tasks to limit external tool access during drafting. Teachers also require students to maintain

complete revision logs recording original outlines draft versions and iterative adjustment records. Short oral defense sessions can be arranged after submission for learners to explain personal writing ideas logical arrangement and cultural expression intentions. Evaluation focus should shift from polished final texts to the whole revision process and students independent thinking tracks. This assessment logic aligns with the core idea of process genre writing teaching which emphasizes learning growth throughout writing cycles (Zhu & Liu, 2024), and is further supported by research on student feedback literacy, which stresses learners' capacity to understand, evaluate, and act upon feedback (Carless & Boud, 2018).The whole set of supervision rules can effectively distinguish independent student creation from fully artificial intelligence generated content and protect the basic construct validity of writing assessment.

3.4.2. Data Compliance for Minor Learners

Long term intelligent evaluation accumulates massive text data including student compositions personal writing weaknesses and learning feature labels. All such records belong to sensitive personal information of minor learners and must be managed under standardized compliance rules. Schools need to follow the personal information protection law and regulations on the protection of minors in cyberspace and stick to the principle of data minimization. Only writing information closely linked to teaching evaluation can be collected without redundant personal labels irrelevant to writing training. Formal informed consent from student guardians should be obtained before opening intelligent evaluation functions. Local school based artificial intelligence platforms are preferred over third party public model products to avoid cross platform data leakage. Clear data storage limits should be set and expired student writing archives need to be permanently erased after graduation. Ethical norms of human artificial intelligence integration provide basic guidance for such data management arrangements and remind educators to balance technical convenience and minor information security (Wen & Tian, 2024).

3.4.3. Balanced Digital Resources and Teacher Artificial Intelligence Literacy Construction

The anytime anywhere evaluation function of large language models assumes stable network environment and sufficient intelligent equipment which cannot be fully guaranteed in all regional schools. Visible gaps in device allocation network stability and authorized platform resources exist between urban and rural teaching environments. Differentiated implementation plans should be formulated for resource limited schools such as concentrated classroom artificial intelligence evaluation sessions and offline printed feedback summaries to narrow the first level digital divide. In addition the human machine collaborative mode puts forward higher demands on teachers professional ability. Regular school based training activities should be held to improve teachers capacity of designing standardized evaluation rubrics judging biased model outputs and guiding student critical use of artificial intelligence feedback. It needs to be clarified that intelligent tools reconstruct teachers work content instead of simply reducing teaching workload. Systematic training frameworks of artificial intelligence assisted foreign language learning literacy offer clear training dimensions for school professional development design (Zhang, 2025). Teacher identity transformation and professional growth in the artificial intelligence era further prove that sustained literacy training is the core foundation for stable promotion of intelligent writing evaluation modes (Xu & Gu, 2025).

4. Limitations and Practical Risks of LLM-Empowered Writing Evaluation

Although large language models have unique auxiliary technical strengths in innovating writing evaluation methods, optimizing evaluation effects and promoting students' writing literacy development, they still have inherent technical defects and practical application limitations in educational teaching scenarios. On the one hand, restricted by training data deviation and algorithm limitations, LLMs have deficiencies in in-depth semantic understanding, innovative logical judgment and emotional expression evaluation, and cannot fully identify students' innovative writing ideas, unique emotional expression and personalized cultural cognition, which may lead to one-sided evaluation results and insufficient evaluation fairness. On the other hand, long-term excessive reliance on intelligent evaluation tools may make students form cognitive path dependence and thinking inertia, lack independent thinking and manual revision training, resulting in superficial learning disengagement and weakening of autonomous writing ability. In addition, the transmission and storage of students' personalized writing data also face potential data security and ethical risks. Over-reliance on intelligent tools will also weaken teachers' humanistic emotional guidance and value leadership in writing teaching, which is not conducive to the cultivation of students' healthy learning attitude and correct cultural values.

https://doi.org/10.67728/sshi.2026.010

5. Research Prospects and Conclusion

5.1. Research Prospects

Continuous iteration of digital intelligence technology will further upgrade the multi dimensional analysis and personalized guidance functions of large language models. Future research can expand along three verifiable empirical directions to enrich the theoretical system of intelligent writing evaluation for senior high school English teaching. First, quasi experimental comparative research can be carried out to track long term changes in student writing literacy. Researchers can set parallel classes adopting pure teacher evaluation and human machine collaborative evaluation separately and record differences in writing accuracy logical expression and cross cultural output ability over a complete semester (Ghafouri et al., 2024; Song & Song, 2023). Second, consistency testing research can compare scoring results from large language models and professional senior high school English teachers. Relevant indicators including exact matching rate and weighted consistency coefficient can be calculated to judge the reliability of model scoring under different writing themes and evaluation rubrics (Guo & Wang, 2024). Third, feedback uptake research can explore factors affecting student acceptance of artificial intelligence revision suggestions. Follow up studies can sort out internal and external factors that drive students to adopt or ignore model feedback and summarize design rules for more effective intelligent evaluation prompts (Wang et al., 2024). In addition, further exploration can focus on differentiated implementation plans for schools with uneven digital resources and long term training systems to improve teachers artificial intelligence literacy. More targeted research can also discuss effective classroom strategies to prevent artificial intelligence academic integrity risks in writing assessment.

5.2. Conclusions

The deep integration of large language model technology and senior high school English writing evaluation brings transformative changes to traditional writing teaching and assessment systems and provides feasible solutions to long standing practical dilemmas including single evaluation dimensions closed evaluation subjects and delayed interactive feedback. Human machine collaborative intelligent evaluation clarifies clear orientation toward disciplinary core literacy and diversified digital intelligent evaluation standards, and constructs three operable teaching paths supported by standardized curriculum requirements visualized learner profiling and cross time space personalized tutoring. Supplementary guarantee measures covering academic integrity minor data compliance balanced digital resources and teacher literacy construction further support the stable classroom application of such evaluation modes.

Intelligent evaluation supported by large language models redefines the role of English teachers from single score judges to whole process learning guides responsible for value judgment and thinking cultivation. With continuous upgrading of educational artificial intelligence algorithms and improvement of school supporting digital infrastructure, human machine collaborative evaluation will become the mainstream mode of senior high school English writing teaching. The whole framework proposed in this research offers theoretical references and operable classroom strategies to promote intelligent refined and developmental transformation of senior high school English writing evaluation and supports high quality progress of basic English education guided by disciplinary core literacy cultivation.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new empirical data were generated or analysed for this conceptual-framework-oriented study. All analyses are built upon publicly-available policy documents, published academic literature and theoretical reasoning. No raw experimental datasets, survey data, interview transcripts or student writing corpora were produced in the course of this research. Should readers require further information regarding the referenced sources, they may access the publicly cited materials listed in the reference section of this article.

Conflict of Interest: The author declares that there exist no competing financial interests, personal connections or other potential conflicts of interest that may affect the research conclusions and manuscript content presented in this article.

https://doi.org/10.67728/sshi.2026.010

References

Bitchener, J. (2012). Written corrective feedback for L2 development. *TESOL Quarterly, 46*(4), 855–860. https://doi.org/10.1002/tesq.62

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal, 63*(2), 97–107. https://doi.org/10.1093/elt/ccn023

Escalante, S., Pack, T., & Barrett, L. (2023).AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, *20*(1), Article 57. https://doi.org/10.1186/s41239-023-00425-2

Ghafouri, M., Hassaskhah, J., & Mahdavi Zafarghandi, A. (2024). From virtual assistant to writing mentor: Exploring the impact of a ChatGPT-based writing instruction protocol on EFL teachers' self-efficacy and learners' writing skill. *Language Teaching Research, 28*(3), 721–742. https://doi.org/10.1177/13621688241239764

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*(7), 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Liu, C. C., Hsieh, S. Y., & Chen, Y. H. (2022). An analysis of children's interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education, 189*, Article 104576. https://doi.org/10.1016/j.compedu.2022.104576

Song, C. P., & Song, Y. P. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*, 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Wang, X. H., Li, Y., & Zhang, L. (2024). Learners' perceived AI presences in AI-supported language learning: A study of AI as a humanized agent from community of inquiry. *Computer Assisted Language Learning, 37*(4), 814–840. https://doi.org/10.1080/09588221.2022.2056203

Zhang, C., Meng, Y. W., & Ma, X. Y. (2024). Artificial intelligence in EFL speaking: Impact on enjoyment, anxiety, and willingness to communicate. *System, 121*, Article 103259. https://doi.org/10.1016/j.system.2024.103259

Cai, J. G. (2025). The use of Artificial intelligence in higher education:An opportunity for the development of studentsindependentcritical thinking. *Foreign Language Teaching and Theory Practice, 193*(1), 61. http://www.teachlanguage.ecnu.edu.cn/EN/Y2025/V193/I1/61 (in Chinese).

Chen, M., & Lü, M. C. (2024). College English Writing Instruction in a ChatGPT Context *Contemporary Foreign Languages Studies, 24*(1), 161–168. https://www.qk.sjtu.edu.cn/cfls/EN/abstract/abstract47760.shtml (in Chinese).

CHEN Y Q, HU J L. (2022). Characteristics of core literacy-oriented curriculum and teaching reform: Based on the interpretation of core-competency traits and learning mechanisms. *Curriculum, Teaching Material and Method*, *42*(10), 12-19. https://www.pep.com.cn/bks/kcjcjf/xszh/202503/t20250313_1998748.shtml

Central Committee of the Communist Party of China, & State Council. (2020). *Overall plan for deepening the reform of educational evaluation in the new era.* Available online: https://www.gov.cn/zhengce/2020-10/13/content_5551032.htm (accessed on 9 September 2026).

Gu, M. Y. (2018). Classroom teaching is the best carrier for thinking cultivation. *Basic Education Forum*, (36), 62–63. http://cnkimirror.clcn.net.cn/KCMS/detail/detail.aspx?filename=JCLT201836025&dbcode=CJFR (in Chinese).

Guo, Q., Feng, R. L., & Hua, Y. F. (2023). Applications and potential risks of ChatGPT in academic English paper writing and teaching. *Media in Foreign Language Education*, (2). https://doi.org/10.20139/j.issn.1001-5795.2023.02.003 (in Chinese).

Hu, X. Y., & Lin, Z. R. (2019). Research on teacher portraits in precision teaching context. *Educational Technology Research, 40*(7), 84–91. https://www.cnki.com.cn/Article/CJFDTOTAL-DHJY201907013.htm (in Chinese).

Kong, L. (2024). Application of generative artificial intelligence in foreign language major teaching: Taking Intensive Reading of Critical Thinking College English as an example. *Frontiers of Foreign Language Education Research*, *7*(1), 11–18, 90. https://doi.org/10.20083/j.cnki.fleic.2024.01.011 (in Chinese).

Li, J. Y., & Zhu, R. Y. (2024). Internal logic and practical innovation of value-added evaluation in Chinese writing teaching. *Curriculum, Teaching Material and Method*, *44*(5), 80–86. https://lib.cqvip.com/Qikan/Article/Detail?id=7112030976 (in Chinese).

Liu, S. J., Zhang, Y., & Huang, L. B. (2025). Translation ethics in the GenAI era: Concept reconstruction, ethical transformation and behavioral choices. *Foreign Language Teaching, 46*(1), 66–72. https://search.napstic.cn/literature/oaj/wyjx?page=2&activeYear=2025&activeIssue=1 (in Chinese).

Liu, Z. J., & Chen, X. C. (2024). Value reflection and reconstruction of teaching evaluation in the digital intelligence era. *Curriculum, Teaching Material and Method*, (10), 67–73. https://mqikan.cqvip.com/Article/ArticleDetail?id=7113442507&from=Article_ArticleDetail (in Chinese).

Lu, Y., Yu, J. L., & Chen, P. H. (2023). Educational application and prospect of generative artificial intelligence: Taking ChatGPT system as an example. *China Distance Education*, *43*(4), 24–31. https://doi.org/10.13541/j.cnki.chinade.20230301.001 (in Chinese).

Ministry of Education of the People's Republic of China. (2020). *General high school English curriculum standards (2017 edition, revised 2020).* People's Education Press.

Sun, J. (2021). Transformation of teaching value in the artificial intelligence era. *Journal of Central China Normal University (Humanities and Social Sciences), 60*(3), 174–181. https://www.scholarmate.com/S/7ryRga (in Chinese).

Wang, S. S., & Wang, J. J. (2024). Motivation-driven research on EFL learners' acceptance of AI-enabled learning based on technology acceptance model. *Media in Foreign Language Education*, (6). https://www.cnki.com.cn/Article/CJFDTotal-WYDH202406004.htm (in Chinese).

Wen, X., & Tian, Y. L. (2024). Integration of human intelligence and artificial intelligence in foreign language education and research. *Media in Foreign Language Education*, (4), 18–24. https://doi.org/10.20139/j.issn.1001-5795.20240403 (in Chinese).

Xu, L. L., Hu, J. H., & Su, Y. (2024). Learner cognition and behavior research on artificial intelligence assisted academic English writing. *Foreign Language World*, (3), 51–58. https://doi.org/10.26971/j.cnki.flw.1004-5112.2024.03.007 (in Chinese).

Xu, C. G. (2025). Research on foreign language learners human machine collaborative writing engagement supported by generative artificial intelligence. *Foreign Languages and Their Teaching*, (2), 61–73. https://d.wanfangdata.com.cn/periodical/wyywyjx202502006 (in Chinese).

Xu, Y. T., & Gu, Y. (2025). Foreign language teaching and teacher professional development in the artificial intelligence age challenges identity crisis and solutions. *Contemporary Foreign Languages Studies, 25*(1), 60–72. https://doi.org/10.3969/j.issn.1674-8921.2025.01.007 (in Chinese).

Zhang, J. (2025). Construction of framework and scale development for university students artificial intelligence empowered foreign language learning literacy. *Media in Foreign Language Education*, *221*(1), 69–78. https://doi.org/10.20139/j.issn.1001-5795.20250110 (in Chinese).

Zhang, Z. Y., & Hong, H. Q. (2023). Foreign language teaching supported by ChatGPT empowering challenges and strategies. *Foreign Language World*, (2), 38–44. (in Chinese).

Zhu, X. C., & Liu, H. N. (2024). Research on application paths of ChatGPT in foreign language process genre writing teaching. *Media in Foreign Language Education*, (6), 30–36. https://d.wanfangdata.com.cn/periodical/wydhjx202406005 (in Chinese).

Zhou, Y. (2017). Research on college English learning environment design from the perspective of ecological affordance. *Foreign Language Teaching*, (6), 70–73. https://doi.org/10.16362/j.cnki.cn61-1023/h.2017.06.013 (in Chinese).

Zheng, S. H., Zhou, D. H., & Zhang, S. H. (2023). Connotation theme reflection and challenges of ChatGPT from computational pedagogy perspective. *Journal of East China Normal University Educational Sciences, 41*, 91–102. https://doi.org/10.16382/j.cnki.1000-5560.2023.07.009 (in Chinese).

Zhang, L. T., & Yang, L. X. (2023). A case study on senior high school English teachers writing evaluation literacy. *Frontiers of Foreign Language Education Research*, *6*(3), 64–72. https://doi.org/10.20083/j.cnki.fleic.2023.03.064 (in Chinese).

Yang, L. R. (2024). Second language acquisition under the background of ChatGPT large language model. *Modern Foreign Languages, 47*(4), 578–585. https://doi.org/10.20071/j.cnki.xdwy.20240523.003 (in Chinese).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Apollo and/or the editor(s). Apollo and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.