ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Research on AI English Writing Feedback Mechanism Based on Deep Semantic Understanding and Its Application in Higher Vocational Teaching

Hui Li[1,*]

[1] School of Fundamental Education, Beijing Polytechnic College, Beijing 100042, China

SUMMARY: *As generative artificial intelligence and natural language processing continue to enter the language teaching scene, how to improve the semantic depth, task adaptability and teaching usability of English writing feedback in higher vocational colleges has become a practical issue in English teaching reform. Focusing on the needs of English writing training in higher vocational colleges, this paper constructs an AI writing feedback mechanism based on deep semantic understanding. Relying on corpus preprocessing, contextual semantic representation, task context matching, error detection, discourse coherence evaluation and multi-task collaborative optimization, this paper realizes a closed-loop design from text input, problem diagnosis, feedback generation and teaching application. The experimental results show that the average precision, recall and F1 value of the model on the error recognition task reach 92.4%, 90.5% and 91.4%, respectively. In terms of feedback content semantic matching, task alignment and suggestion relevance, the proposed method achieves 89.6, 87.9 and 90.8, respectively. In the teaching application, the total score of the post-test in the experimental group increased from 70.5 to 81.7, and the acceptance rate of students 'feedback increased from 68.4% to 84.6%. The research shows that this method can effectively improve the feedback quality and revision efficiency, and has practical significance for promoting the intelligent and refined development of English writing teaching in higher vocational colleges.*

KEYWORDS: *Deep semantic understanding; AI English writing; Feedback mechanism; Higher Vocational teaching application*

1 Introduction

In recent years, generative artificial intelligence, natural language processing and educational big data technology continue to enter the language teaching scene, and English writing teaching has gradually shifted from traditional outcome evaluation to process support and intelligent intervention. For higher vocational colleges, English writing is not only related to students 'mastery of language knowledge, but also directly affects their professional communication, job expression and cross-cultural communication ability. Compared with ordinary undergraduate colleges, the English foundation of higher vocational college students is more different, and problems such as single use of vocabulary, unstable syntactic structure, loose cohesion of discourse and unclear expression purpose often appear in writing training. Through teaching observation and analysis of writing training, we found that teachers often need to take into account grammar error correction, semantic discrimination, discourse organization analysis and personalized suggestion generation in the process of marking. The task is heavy and the

*amier_Lee@163.com
https://doi.org/10.65102/is2026223 Published: 30 April, 2026

feedback cycle is long, which is difficult to continuously meet the practical needs of students 'repeated practice and instant correction. How to use artificial intelligence technology to improve the accuracy, pertinency and teaching adaptability of writing feedback has become an important issue in the reform of English teaching in higher vocational colleges.

The existing intelligent assessment tools of English writing have improved the efficiency of assessment to a certain extent, but many of them still stay at the stage of surface feature recognition, and pay more attention to explicit indicators such as spelling, grammar, sentence length and word frequency. The recognition ability of contextual semantic relationship, writing intention, discourse coherence and pragmatic relevance is still insufficient. We believe that although part of the feedback content can point out the location of the error, it is difficult to explain the cause of the problem, and it is difficult to give modification suggestions that match the cognitive level of students. This kind of feedback method focuses on formal error correction and weakens deep understanding, which is easy to make students understand writing training as partial sentence repair, and it is difficult to truly improve the quality of content organization and expression. Especially in higher vocational English teaching, writing tasks are often closely related to professional context, applied style and practical expression. If the feedback mechanism cannot combine with specific teaching objectives, it is difficult to play a stable role in teaching promotion.

The development of deep semantic understanding technology provides a new way to improve this problem. Relying on the methods of pre-trained language model, contextual representation learning, text semantic matching and multi-task learning, the system can analyze student texts in more detail from the level of vocabulary, sentence and discourse, and understand the deep problems such as semantic deviation, logical break and expression imbalance while identifying language errors. We believe that if the writing feedback mechanism is further combined with the teaching process in higher vocational colleges, and a complete link covering text input, semantic representation, problem diagnosis, feedback generation and teaching application is constructed, it is possible to realize the transformation from a single scoring tool to an intelligent teaching support system.

Based on this, this paper focuses on the real needs of English writing teaching in higher vocational colleges, studies the AI English writing feedback mechanism based on deep semantic understanding, and discusses its application in higher vocational teaching. This paper will analyze the writing corpus preprocessing, deep semantic feature construction, feedback generation mechanism, teaching implementation process and experimental evaluation. Through this research, we hope to improve the semantic accuracy and teaching usability of writing feedback, and provide more intelligent, refined and continuous technical support for English writing teaching in higher vocational colleges.

2 Related work

The development of automatic writing assessment and intelligent feedback technology provides a new support path for English writing teaching. Early related studies paid more attention to how teachers introduced automatic writing evaluation systems into the classroom, and the influence of system feedback on teaching implementation. Li pointed out that with the support of AWE system, teachers' cognition, adoption mode and classroom organization strategy of automatic feedback would directly affect the play of feedback effect, indicating that intelligent writing tools are not technical modules independent of teaching situations, but need to participate in the writing guidance process together with teachers' judgment [1]. Ranalli further investigated the participation of second language students in automatic feedback from the perspective of learners, and found that whether students really use feedback to revise is closely related to their judgments on feedback credibility, interpretability and learning value, indicating that the effectiveness of feedback depends not only on whether the system can identify errors, but also on whether students can understand and trust the feedback content [2]. Hassanzadeh and Fotoohnejad's research in foreign language writing courses showed that learner-centered automatic feedback application could improve the convenience of revision and the frequency of practice, but its promoting effect was still restricted by the design of learning tasks, the way of feedback presentation and students' self-revision ability [3]. From this, we can see that although the automatic writing evaluation technology improves the feedback efficiency, its real role cannot be separated from the support of the specific teaching situation.

With the accumulation of research, scholars have begun to systematically sort out the application effects of automatic writing evaluation systems from the review level. After reviewing the relevant studies in the school field from 2000 to 2020, Nunes et al. pointed out that AWE system has obvious advantages in improving the frequency of writing exercises, reducing the burden of teachers' correction and supporting immediate feedback, but there is still a lack of sufficient evidence in the depth of feedback, the difference of applicable objects and the long-term transfer effect [4]. Guo et al. focused on the use of automatic written error correction feedback in EFL students' research writing, and found that students could improve some language form problems with the help of the system, but the support of automatic feedback was relatively limited for revision tasks requiring strong semantic inference and discourse integration [5]. Yu et al. proposed and verified a writing feedback literacy scale for Second language students, indicating that students' ability to understand, screen, absorb and apply feedback itself is an important variable affecting the effectiveness of feedback, and suggesting that the research on intelligent feedback should not only focus on system performance, but also focus on the cultivation of learners' feedback processing ability [6]. We believe that this stage of research has extended the focus from the tool itself to the learner's use process, which provides an important foundation for subsequent research from the perspective of deep semantics and instructional adaptation.

On this basis, the research perspective gradually shifts from "whether the system is effective" to "why the system is effective and under what conditions". Li reexamined the role of automatic writing evaluation tools from the perspective of activity theory and pointed out that AWE is still an auxiliary tool with limitations, and its use effect is affected by multiple factors such as task objectives, classroom rules, division of labor between teachers and students, and evaluation culture, so the improvement of writing cannot be simply attributed to technology itself [7]. Shi and Aryadoust systematically reviewed the research on AI-driven automatic written feedback. According to Shi and Aryadoust, stronger artificial intelligence methods have been introduced in recent years, but the existing systems still generally emphasize on the recognition of lexical, grammatical and local revised levels, and the support for contextual semantic relations, argumency logic and text level feedback is still insufficient [8]. Karatay and Karatay also pointed out after synthesizing the AWE research in second language classroom that the role of automatic assessment tools in different learning stages and different teaching scenarios is not balanced, and how to achieve the synergy between technical feedback and teacher feedback, classroom tasks and learning objectives is still an important issue in current research [9]. Sari and Han further showed that automatic writing evaluation not only affected students' writing performance, but also had a linkage effect on writing self-efficacy, self-regulation and anxiety level, indicating that intelligent feedback mechanism was no longer a simple error correction tool, but was becoming an important teaching variable affecting the learning process [10]. After combing the AWE feedback research from three levels of feedback types, learning results and teaching enlightenment, Fu et al. pointed out that more attention should be paid to the hierarchy, interpretation and operability of feedback content in the future, so as to promote feedback from single error correction to more diagnostic and developmental writing support [11]. Combined with the above studies, we can find that the current development direction of automatic writing evaluation is gradually extending from surface error correction to deep understanding, and from single output to teaching collaboration.

In general, the existing research lays a good foundation for automatic writing evaluation and intelligent feedback teaching applications. It not only proves the value of AWE in improving feedback efficiency and supporting multiple rounds of revision, but also reveals its shortcomings in feedback credibility, semantic depth, discourse diagnosis and teaching adaptability. We believe that, especially in the scenario of English writing in higher vocational colleges, students' writing tasks are often connected with professional communication, application expression and situational practice, and it is difficult to meet the actual needs of writing ability improvement by relying solely on surface language features for feedback. Therefore, it is of strong research necessity and practical value to construct an AI English writing feedback mechanism with more explanatory, targeted and teaching adaptability around deep semantic understanding, and embed it into the teaching implementation process of higher vocational colleges.

3 The AI English Writing Feedback Mechanism Based on deep Semantic Understanding and the Design of Higher Vocational Teaching Application

3.1 Preprocessing of Higher Vocational English Writing Corpus and Construction of Deep semantic features

In order to give stable, interpretable and teaching-specific suggestions, the premise of the English writing feedback system in higher vocational colleges is not only to "read the text", but also to transform the vocabulary selection, syntactic organization, semantic expression and task goal of students 'writing into computable and comparable feature representations at the same time. The writing texts of higher vocational students often have obvious level differences. Some texts have surface errors such as spelling, tense and collocation, and the other part of the problems are reflected in the deeper position such as unstable intersentence cohesion, broken semantic advancement, fuzzy information focus and insufficient sense of genre. If the system only relies on keyword matching or local grammar rules, it is easy to leave the feedback at the level of mechanical error correction, which is difficult to support the subsequent deep feedback generation. Therefore, in the early stage of system design, it is necessary to normalize the writing corpus, and then establish the deep semantic feature space oriented to word, sentence and article, so that the feedback mechanism has the comprehensive perception ability of text content, context relationship and task context.

In the corpus preprocessing stage, the system uniformly transformed the essays submitted by students, the time-limited writing texts in class, the training samples of applied essays and the revised drafts into standard text sequences. Firstly, the original text is denoised, including abnormal symbol cleaning, case normalization, abbreviation expansion and clause segmentation. Then, combined with part-of-speech tagging, dependency relationship analysis and named entity recognition, the text is structured. Considering that there are many transfer traces between interlanguage expressions and Chinese in English writing in higher vocational colleges, the system also needs to retain some "non-standard but diagnostic value" expression fragments to avoid smoothing out meaningful teaching problems in advance in the preprocessing stage. After processing, the text is divided into word sequence, sentence sequence and discourse sequence to provide a unified input for subsequent semantic modeling.

In the deep semantic representation stage, we use context encoding to generate word-level semantic vectors. Let the word vector of the student composition at time t be xt, and its context hidden state be represented as follows.

$$h_t = Encoder(x_t, h_{t-1}) \tag{1}$$

where, h_t represents the semantic state of the current word in the context, and h_{t-1} represents the context information at the previous position. In order to avoid too scattered word-level features, the system further aggregates the internal information of the sentence to construct the sentence-level semantic representation. Let a sentence s_i consist of n context vectors, then the sentence vector can be expressed as follows.

$$u_i = \frac{1}{n}\sum_{t=1}^{n} \alpha_t h_t \tag{2}$$

where α_t is the semantic weight of the term in the current sentence. Different from simple average, this weighted aggregation method can highlight central predicates, logical connectives, evaluative expressions and task-critical information, so that the system is closer to the focus of teachers' marking when analyzing sentential semantic integrity. Based on the set of sentence vectors, the system combines paragraph position, cohesion markers and topic advancement information to form a discourse level representation, which is used to identify content repetition, discourse jump, and loose discourse.

Considering that the feedback of English writing in higher vocational colleges should not only judge whether the writing is correct, but also judge whether it meets the task requirements, the system introduces the task context matching mechanism in the feature construction. Let discourse semantics be represented by d and writing task goals be represented by g, then the matching score between text and task is defined as follows.

$$r = \frac{d \cdot g}{|d||g|} \tag{3}$$

where, r reflects the closeness between the student text and the target task at the semantic level. If the value is low, it means that although students may not make obvious mistakes at the grammar level, they still deviate from the content direction, genre requirements or expression purpose. This index can provide a basis for the subsequent feedback module, so that the system can be further extended from "whether the language is correct" to "whether the expression is meaningful and whether it is synthesized".Figure 1 shows the overall framework of AI English writing feedback driven by deep semantic understanding.

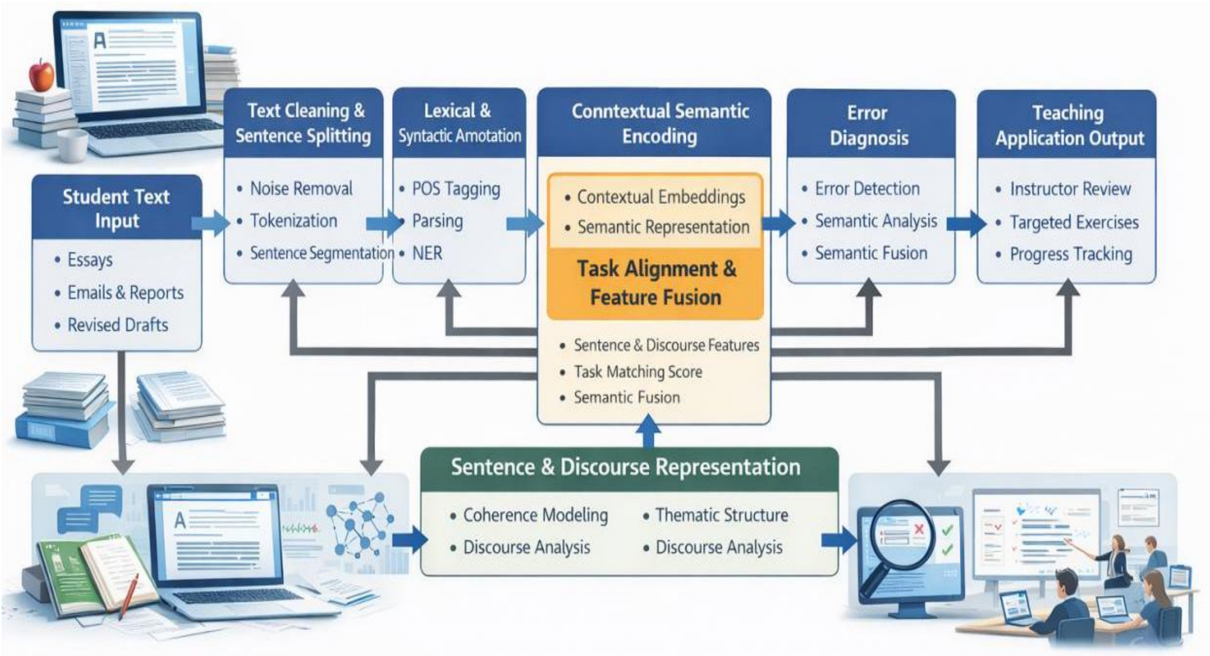

Figure 1: General framework diagram of AI English writing feedback driven by deep semantic understanding

In order to enhance the consistency and computability of system input, different corpus types adopt a unified specification in preprocessing and feature output, as shown in Table 1.

Table 1: Corpus preprocessing and composition of deep semantic features in Higher Vocational English writing

Corpus Type	Input Content	Preprocessing Operations	Output Features
Timed Classroom Essays	Students' one-draft writing texts	Noise removal, sentence segmentation, part-of-speech tagging	Lexical density, syntactic structure, tense distribution
Practical Writing	Emails, notices, cover letters, etc.	Genre segmentation, format recognition, dependency parsing	Genre elements, sentence pattern templates, semantic intent
Process-oriented Revised Drafts	Initial drafts and revised drafts	Version alignment, difference annotation	Revision traces, problem retention rate, revision depth
Occupational-context Writing	Workplace communication and scenario-based expression texts	Keyword extraction, named entity recognition, context mapping	Contextual vocabulary, functional expressions, content matching degree
Reference Model Corpus	Teaching samples and excellent essays	Sentence-group segmentation, discourse structure annotation	Thematic progression, cohesion patterns, expression paradigms

Through the above processing, the system completed the transformation from original student text to deep semantic features, which laid a computable, interpretable and teaching-adaptable input foundation for the subsequent feedback generation mechanism. On this basis, Section 3.2 will further discuss the generation logic of AI English writing feedback and the multi-task collaborative modeling method.

3.2 Collaborative Modeling of feedback Generation Mechanism and feedback task in AI English Writing

After the preprocessing of writing corpus and the construction of deep semantic features, the system needs to further answer two core questions. First, what kind of questions appear in the student text? Secondly, in what way these questions should be transformed into understandable, executable and suitable feedback content for English teaching in higher vocational colleges. Different from the traditional automatic correction system which mainly stops at spelling, grammar and fixed collocation recognition, we divide the feedback generation into four continuous steps: error detection, semantic diagnosis, discourse evaluation and feedback decision, so that the system can not only find local errors, but also identify deep problems such as content deviation, logical break and task mismatch. In the lexical and syntactic error recognition stage, the system first classifies and judges the fused representation z_i of the ith writing unit, and obtains the probability that the unit belongs to the KTH error type:

$$p_{i,k} = \frac{\exp(W_k z_i + b_k)}{\sum_{c=1}^{K} \exp(W_c z_i + b_c)} \tag{4}$$

where K represents the total number of error categories, W_k and b_k are classification parameters. This formula is used to distinguish different types of errors such as spelling errors, tense errors, subject-verb agreement errors, collocation anomalies and sentence pattern defects. Compared with the single binary classification method, this multi-category decision is more suitable for the fine-grained invocation of the subsequent feedback template, because different errors correspond to different explanation strategies and modification suggestions.

It is difficult to support high-quality feedback by only relying on local bias detection. Some students' sentences have no obvious problems on the surface of grammar, but the expression center is shifted, the semantic information is incomplete or disconnected from the requirements of the topic. To this end, the system introduces the semantic deviation discrimination of sentence segments. Let the semantic vector of the JTH sentence segment be s_j and the task reference semantic center be t_j, then the semantic deviation score is defined as follows.

$$d_j = \|s_j - t_j\|_2^2 \tag{5}$$

When d_j is large, it indicates that the paragraph has deviated from the writing task in content focus, semantic orientation or information expression, although it may have the correct form. The index can help the system identify the questions "written smoothly but not answered correctly", and make the feedback from the language surface error correction further into the semantic level diagnosis.

In order to enhance the grasp of discourse structure issues, the system continues to estimate the coherence relation between adjacent sentence segments. Let the representations of paragraph j and paragraph j+1 be u_j and u_{j+1}, respectively, then the local coherence score is written as follows.

$$c_j = \frac{u_j^T u_{j+1}}{\|u_j\| \|u_{j+1}\|} \tag{6}$$

Furthermore, the discourse coherence of the whole essay can be expressed as follows.

$$C = \frac{1}{m-1}\sum_{j=1}^{m-1} c_j \tag{7}$$

where, m is the number of sentence segments. This group of formulas is mainly used to determine whether there are problems such as inter-sentence jump, unstable topic advancement and loose paragraph cohesion in the student text. English writing in higher vocational colleges emphasizes the clarity and completeness of application expression, so discourse coherence evaluation is an indispensable part of feedback mechanism.

After obtaining the information of local bias, semantic bias and discourse coherence, the system needs to map the multi-source diagnosis results into the final feedback strength uniformly. Let the standardized results of bias risk, semantic bias and discourse quality be e_i, d_i and q_i, respectively, then the comprehensive feedback weight of the ith unit can be defined as follows.

$$\alpha_i = \frac{\exp(\gamma_1 e_i + \gamma_2 d_i + \gamma_3 q_i)}{\sum_{r=1}^{n}\exp(\gamma_1 e_r + \gamma_2 d_r + \gamma_3 q_r)} \tag{8}$$

Here, γ_1, γ_2, and γ_3 are the adjustment coefficients for different diagnostic dimensions. In the feedback text generation stage, the system inputs the comprehensive feature vector g_i into the generation module to form the candidate feedback representation:

$$r_i = \tanh(W_r g_i + b_r) \tag{9}$$

The feedback confidence is then combined to decide whether to output the suggestion directly. The confidence is calculated as follows.

$$\eta_i = \sigma(w_\eta^T r_i + b_\eta) \tag{10}$$

When η_i is higher than the preset threshold, the system pushes the direct modification suggestion to the student end. When η_i is in the middle interval, the explanatory prompt is preferentially output or the question is marked to the teacher review end. This can reduce the misdirection caused by low reliable feedback to students, and also make the system more in line with the actual logic of "machine screening - teacher checking" in classroom teaching.

In order to optimize different feedback tasks collaboratively in the same framework, this paper constructs a multi-task joint objective function:

$$L = \lambda_1 L_{cls} + \lambda_2 L_{sem} + \lambda_3 L_{coh} + \lambda_4 L_{gen} \tag{11}$$

Here, L_{cls} represents the bias classification loss, L_{sem} represents the semantic bias constraint loss, L_{coh} represents the discourse coherence evaluation loss, L_{gen} represents the feedback generation loss, and λ_1 to λ_4 are the corresponding task weights. In the process of joint optimization, the bias classification task is mainly responsible for locating the explicit problems at the level of vocabulary, grammar and sentence pattern, and providing the basic diagnosis results for the system. The semantic deviation constraint task focused on judging whether students' expressions deviated from the topic intention, context requirements and content focus, and improved the system's ability to identify deep semantic imbalance problems. The discourse coherence assessment task is used to capture potential defects in inter-sentence cohesion, paragraph advancement and overall structure organization, so that the feedback is not limited to local sentence repair. The feedback generation task completes the reorganization and language organization of suggestions according to the above diagnostic information, so as to ensure the readability, pertinence and operability of the output content. The co-training of multiple subtasks under a unified objective function can reduce the information bias caused by the dominance of a single task, and make the model form a more stable association mapping between different levels.

3.3 The application process and Teaching implementation Design of writing feedback for English teaching in Higher Vocational Colleges

English writing teaching in higher vocational colleges emphasizes task-orientation, situational application and continuous improvement. Therefore, if the AI feedback system really enters the classroom, it should not only complete text analysis and error warning, but also embed a complete teaching link of "task release - first draft writing - intelligent feedback - student revision - teacher intervention - classroom reinforcement - resubmission". Different from general automatic correction, higher vocational teaching pays more attention to whether students can achieve effective output in practical writing, career communication expression and job scene simulation, which requires that the feedback mechanism not only serves the single error correction, but also serves the process training and stage improvement. Based on this, we design the writing feedback application process as a closed-loop structure with the participation of the student end, the system end and the teacher end, so that the feedback results can continue to play a role in classroom, after-class and stage evaluation.

In the early stage of teaching implementation, teachers issued writing tasks according to unit goals, and set genre requirements, situational conditions and evaluation points simultaneously. After students complete the first draft, the system performs deep semantic analysis of the text, and generates hierarchical feedback from four levels: vocabulary accuracy, sentential semantic integrity, discourse coherence and task matching. In order to measure the degree of students' absorption of system feedback, suppose that the confidence of the feedback of article j received by the ith student is q_{ij}, and whether it is adopted is $m_{ij} \in \{0,1\}$, then the feedback adoption rate can be defined as follows.

$$A_i = \frac{\sum_{j=1}^{n_i} q_{ij}\, m_{ij}}{\sum_{j=1}^{n_i} q_{ij}} \tag{12}$$

where n_i represents the total number of feedback received by the student. This formula reflects students' actual use of effective feedback, and can avoid treating low confidence or marginal suggestions as the same as critical feedback, so as to evaluate students' revision participation more truly.

After the student completes the first revision based on the feedback, the system continues to compare the quality change between the first draft and the revised draft to determine whether the feedback actually translates into writing improvement. Let the improvement of language accuracy, improvement of semantic expression, improvement of discourse organization and improvement of task completion be Δa_i, Δs_i, Δc_i and Δt_i, respectively, then the revision gain score of the ith student is as follows.

$$G_i = \beta_1 \Delta a_i + \beta_2 \Delta s_i + \beta_3 \Delta c_i + \beta_4 \Delta t_i \tag{13}$$

Here, β_1 to β_4 are the weights of each dimension. This index can distinguish "how much has been revised" from "how much has been improved", so that the system not only pays attention to the surface revision behavior, but also pays attention to the actual quality improvement after revision, which is consistent with the requirements of effective results and applicable expression in English writing teaching in higher vocational colleges.

In the teacher intervention phase, the system automatically generated intervention priorities based on feedback adoption, revision gain and problem severity. Let the severity of the student's current problem be Ei, then the teacher's intervention priority is defined as follows.

$$P_i = \mu_1(1 - A_i) + \mu_2(1 - G_i) + \mu_3 E_i \tag{14}$$

where, a higher value of P_i indicates that students need teacher involvement more in terms of feedback absorption, revision effect, or problem complexity. Based on this value, teachers can give priority to high-risk students, focus on explaining common problems, and carry out targeted guidance for students with low adoption rate, so as to improve the use efficiency of classroom time.

In order to support teaching decisions at the class level, the system further aggregates and analyzes the common problems in the same task. Let the frequency of the KTH question in the ith student's essay be f_{ik} and the average severity be w_{ik}, then the class common question index can

It is expressed as:

$$R_k = \frac{1}{N}\sum_{i=1}^{N} f_{ik} w_{ik} \tag{15}$$

where, N is the total number of students in the class. Through this formula, the system can identify the key problems that are most worthy of classroom evaluation in a certain round of writing, such as frequent errors in genre format, obvious deviation of task direction, or general weakness of paragraph cohesion, so as to help teachers further transform individual feedback into class teaching content.

After the completion of class evaluation and special training, the system also needs to dynamically update students' writing ability. Let the student's writing ability state after the ith training be $M_i^{(t)}$, then its ability estimation before the next training round can be written as follows.

$$M_i^{(t+1)} = M_i^{(t)} + \rho G_i - \delta E_i \tag{16}$$

Here, ρ represents the promotion coefficient of revision gain on capability improvement, and δ represents the inhibition coefficient of residual problem on capability development. The significance of this formula is that the single writing training is incorporated into a continuous learning trajectory, so that the system can recommend subsequent training tasks, adjust the feedback granularity, and form a more personalized teaching support path.The application process of writing feedback for English teaching in higher vocational colleges is shown in Figure 2.

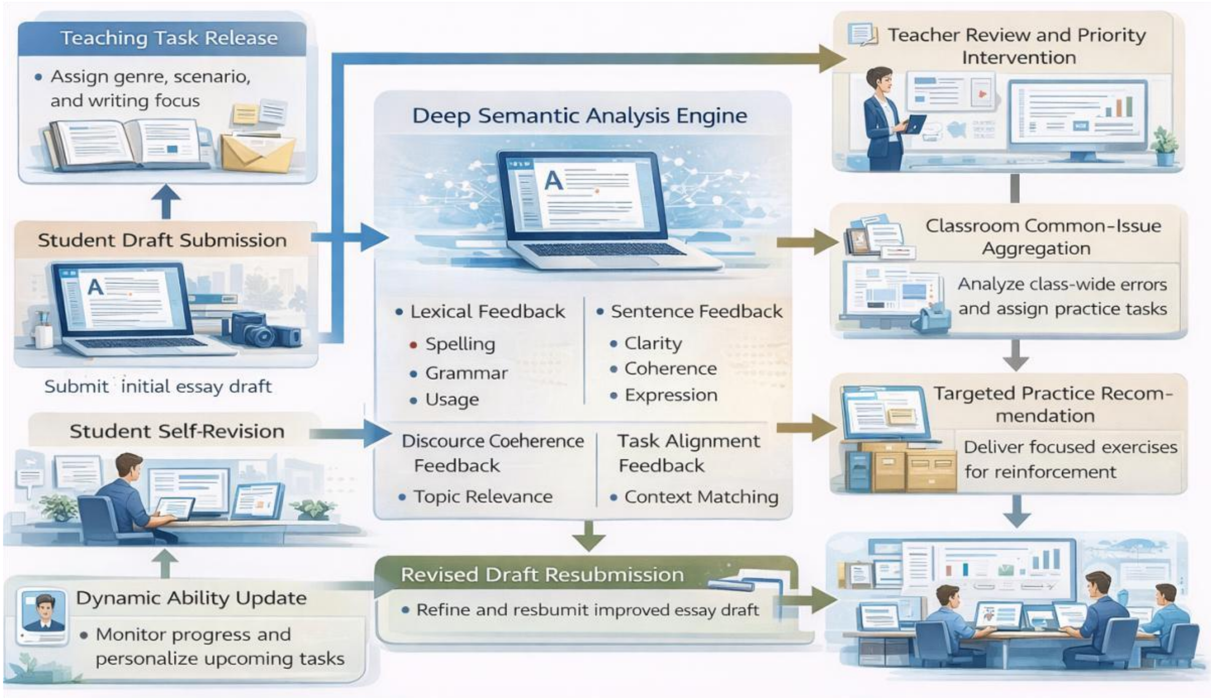

Figure 2: The flowchart of the application of writing feedback for English teaching in higher vocational colleges

As shown in Figure 2, the application process of writing feedback for English teaching in higher vocational colleges is not a linear correction process, but a closed teaching loop with multi-agent participation and cyclic iteration. The teacher first issues the writing task and sets the evaluation focus, and the students enter the system analysis stage after completing the first draft. After the hierarchical feedback output from the system, students made independent revisions, and teachers reviewed and directed according to the intervention priority. Then, the classroom analysis module aggregated the common problems to generate targeted reinforcement training content. Students continue to revise in the second training or submission, and the system synchronously records their learning trajectory and ability changes. Through this process, AI feedback becomes an important supporting mechanism that connects writing training, class reviews, and continuous improvement.

4 Experimental Settings and teaching application environment configuration

4.1 Experimental environment and system operation configuration

In order to ensure that the AI English writing feedback model based on deep semantic understanding can stably complete tasks such as text encoding, semantic matching, bias identification and feedback generation, we construct a relatively complete experimental environment and system operating platform. In the hardware part, the high-performance computing server is used as the main carrier, and the central processor is Intel Xeon Silver 4314, which runs at 2.40 GHz and has 32 computing cores. The graphics processing unit used two NVIDIA RTX 3090 graphics cards with 24 GB memory per card. System memory is configured to 128 GB DDR4. The storage device uses 2 TB NVMe SSD to meet the operational requirements of large-scale text reading and writing, model parameter storage, and intermediate result caching. This configuration can well support deep semantic representation learning and parallel computing of multi-task feedback model.

In terms of software environment, the system runs on Ubuntu 20.04 LTS operating system, the development language is Python 3.10, and the deep learning framework is PyTorch 2.1. The main tools for text preprocessing and data analysis include NumPy 1.26, Pandas 2.0, and spaCy 3.7. The database management system uses MySQL 8.0, which is used to save the essay text, semantic feature index, feedback results and log information. The front-end interactive interface is deployed based on the Flask framework to realize the basic functions such as essay upload, feedback display and result call. The overall software environment has good compatibility and scalability, and can meet the unified operation requirements of model training, online reasoning and result management.

In the system operation parameter Settings, the model training batch size is set to 32, the maximum input sequence length is set to 512 tokens, the initial learning rate is set to 0.0005, the optimizer uses AdamW, the weight decay coefficient is set to 0.01, the number of training rounds is set to 100, and the number of early stopping rounds is set to 8. In order to improve the efficiency of online feedback, the system adopts the operation mode of "offline training + online reasoning", in which the offline part is responsible for model training and parameter updating, and the online part is responsible for student composition reception, semantic analysis, feedback generation and result return. On the whole, the experimental environment and operating configuration can meet the basic requirements of model training and system deployment in this paper, and provide stable technical support for subsequent experimental analysis.

4.2 Data set construction and teaching experiment process design

In order to ensure that the experimental results are comparable and application-specific, we construct a teaching experiment dataset for English writing feedback tasks in higher vocational colleges. The data mainly came from classroom time-limited writing, after-class practical writing and stage revision, and anonymization processing, text cleaning, clause labeling and task label coding were completed uniformly. After comprehensive collation, there were 1260 original essays and 840 revised essays, with a total sample size of 2100. The writing tasks covered types such as email writing, notice writing, job application and report writing. The dataset is divided into training set, validation set and test set according to the ratio of 7:1.5:1.5 to ensure that the model training, parameter adjustment and effect evaluation are independent of each other.

The teaching experiment was carried out by the process of "first draft writing-system feedback-student revise-teacher review-result record". After students completed the first draft, the system conducted a comprehensive analysis around vocabulary error, sentential semantic deviation, discourse coherence and task matching degree, and generated hierarchical feedback results. Students complete the revision according to the feedback, and the teacher makes a targeted review based on the revision performance. Finally, the experiment comprehensively evaluates the feedback effect from four dimensions of language accuracy, semantic expression, discourse coherence and task completion. Combined with the change of student feedback acceptance rate and revision gain, the practical application value of the constructed model in English writing teaching in higher vocational colleges is further tested.

5 Result analysis and evaluation of teaching application effect

5.1 Writing error identification and feedback accuracy analysis

In order to test the error recognition ability and feedback accuracy of the constructed model in higher vocational English writing scenarios, this paper selected typical essay samples from the test set, classified and statistically analyzed six kinds of high-frequency problems, including spelling, tense, subject-verb agreement, collocation, sentence structure and punctuation format, and evaluated them from three dimensions of accuracy, recall rate and F1 value. The relevant results are shown in Table 2.

Table 2: Comparison of recognition effects for different error types

Error Type	Number of Samples	Precision / %	Recall / %	F1 Score / %
Spelling Errors	186	95.4	93.8	94.6
Tense Errors	172	92.7	90.8	91.7
Subject–Verb Agreement Errors	158	91.6	89.9	90.7
Collocation Errors	149	89.8	87.5	88.6
Sentence Structure Errors	137	88.9	86.7	87.8
Punctuation and Formatting Errors	121	96.1	94.5	95.3
Overall Average	923	92.4	90.5	91.4

As can be seen from Table 2, the system performs better in identifying error types with clear surface rules, in which the accuracy of punctuation and format errors reaches 96.1%, the F1 value reaches 95.3%, and the precision and recall rate of spelling errors reach 95.4% and 93.8%, respectively, indicating that the model has strong stability in character-level feature capture and rule constraint judgment. In contrast, collocation errors and sentence structure errors are more difficult to identify, with F1 scores of 88.6% and 87.8%, respectively. Such problems are often closely related to context semantics, syntactic combination and expression habits, and a single local information is difficult to completely cover their changes. In general, the average precision of the model on 923 error samples reaches 92.4%, the average recall rate is 90.5%, and the comprehensive F1 value is 91.4%. It shows that the feedback mechanism can accurately locate the main problems in higher vocational English writing, and provide a reliable basis for subsequent semantic matching analysis and teaching application evaluation.

5.2 Analysis of semantic matching effect of feedback content

In the process of English writing feedback, the system should not only point out language problems, but also ensure that the feedback content is consistent with the semantics of the original text, the goal of the writing task and the revision needs of the students. To this end, this paper further compares and analyzes the content quality of different feedback methods from three dimensions: semantic matching score, task alignment score and suggestion relevance, and the specific results are shown in Figure 3.

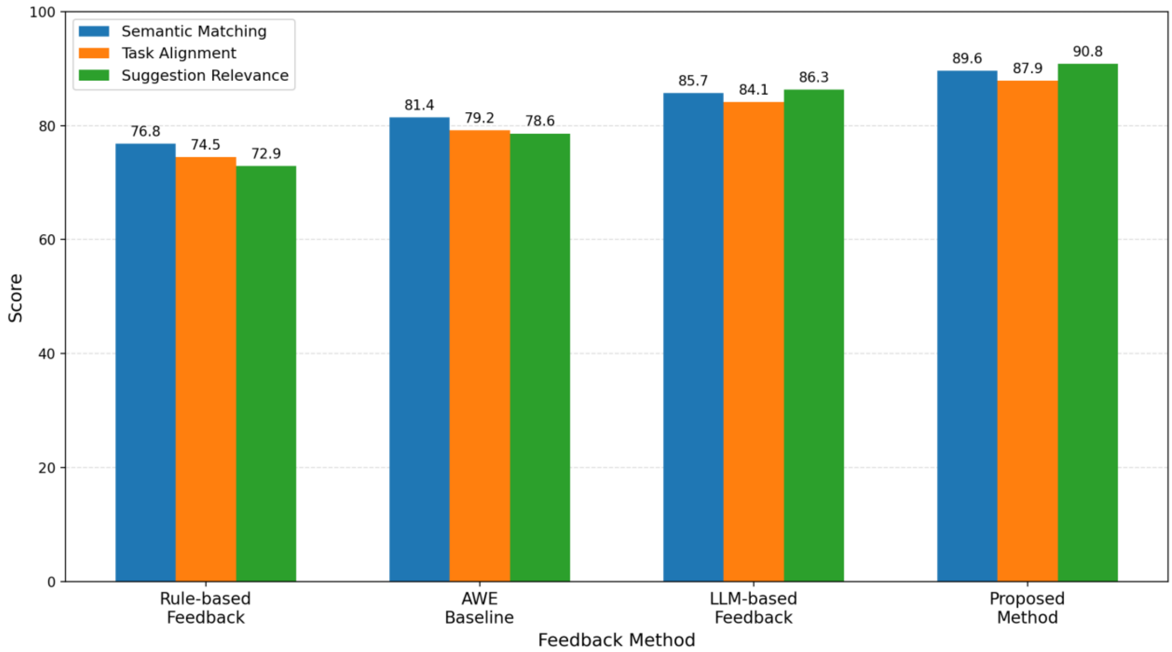

Figure 3: Comparison of feedback content semantic matching effect

As shown in Figure 3, the semantic matching score, task alignment score and suggestion relevance score of rule template feedback are 76.8, 74.5 and 72.9, respectively. The overall performance is relatively stable, but the adaptability to complex context and deep expression problems is insufficient. The corresponding scores of traditional AWE feedback are 81.4, 79.2 and 78.6, respectively, indicating that it has certain advantages in basic language problem recognition, but still has limitations in content layer interpretation and task adaptation. The three scores of the general large model feedback are improved to 85.7, 84.1, and 86.3, showing strong context understanding ability, but the stability and pertinence of its output results in teaching scenarios are still fluctuating. In contrast, the proposed method achieves 89.6, 87.9, and 90.8 in three indicators, respectively, which are the highest in each group.

This result indicates that the feedback mechanism constructed based on deep semantic understanding is more effective in establishing the correspondence between student texts, task requirements, and feedback suggestions. First, the system introduces the process of semantic deviation diagnosis and task intention matching before feedback generation, so that the suggestion output no longer depends on a single local feature, but combines the context and the writing goal to make a comprehensive judgment. Secondly, the multi-task collaborative modeling enables the model to learn bias recognition, semantic interpretation and feedback organization synchronously in the training phase, so that the generated suggestions are closer to the teacher's manual feedback in content integrity and expression specificity.

5.3 Analysis of the contribution of different feedback characteristics to the writing evaluation results

In order to further analyze the actual role of various feedback features in writing evaluation, we use the ablation contrast method to test the contributions of lexical features, syntactic features, semantic features and discourse features. Under the condition of keeping the data set, training parameters and evaluation criteria consistent, the model is trained again after removing a certain type of features respectively, and the performance differences of the system under different configurations in the comprehensive evaluation task are compared. The relevant results are shown in Table 3.

Table 3: Effects of different combinations of feedback features on writing evaluation results

Model Configuration	Writing Evaluation Accuracy / %	Semantic Matching Score	Task Completion Score	Overall Feedback Quality Score
Without Lexical Features	86.9	84.1	82.7	84.5
Without Syntactic Features	85.8	83.6	82.1	83.7
Without Semantic Features	82.4	79.8	78.5	80.2
Without Discourse Features	84.1	81.7	80.9	81.8
Full Model	89.7	87.9	86.8	88.6

It can be seen from Table 3 that the complete model achieves the best results in all indicators, with the writing evaluation accuracy of 89.7%, the semantic matching score of 87.9, the task completion score of 86.8, and the comprehensive feedback quality score of 88.6. In contrast, the performance of removing semantic features decreases most obviously, the accuracy of writing evaluation decreases to 82.4%, and the comprehensive feedback quality score decreases to 80.2, which is 7.3 and 8.4 points lower than that of the full model, respectively, indicating that semantic features play a key role in judging whether the feedback conforms to students 'expression intention, task requirements and modification direction. After removing the discourse features, the degradation of the system's task completion and comprehensive feedback quality is also more prominent, indicating that the discourse level information plays an important supporting role in evaluating the overall structure and logical coherence of the text. Although the lexical features and syntactic features are more basic, they still directly affect the judgment of language accuracy and the generation of local feedback. In general, the four types of features are not independent of each other in writing evaluation, but together constitute the feedback basis from local error correction to overall diagnosis, and the contribution of semantic features and discourse features is more significant.

5.4 Model training convergence and system stability analysis

In order to verify the convergence characteristics and operation stability of the constructed model in the training process, this paper further calculates the training loss, validation loss and validation accuracy under the change of training rounds, and analyzes the fitting state and generalization trend of the model in the process of parameter updating. The relevant results are shown in Figure 4.

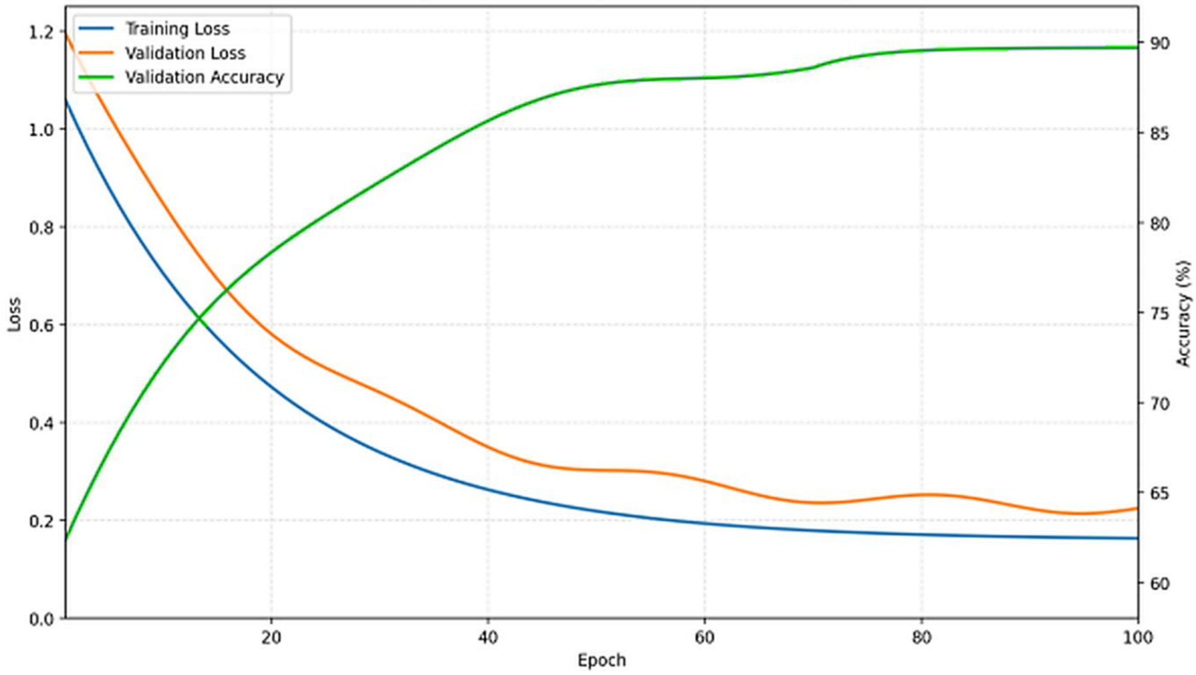

Figure 4: Curve of model training convergence and system stability change

As shown in Figure 4, in the early stage of training, the model parameters are in the stage of rapid adjustment, and both training loss and validation loss show a significant downward trend, indicating that the deep semantic representation and multi-task feedback mechanism can quickly capture the main features in the written text. With the increase of training rounds, the training loss continues to decrease from 1.06 in the initial stage to 0.17, and the validation loss gradually decreases from 1.19 to 0.24. The fluctuation amplitude of both of them is significantly reduced in the later stage, indicating that the model has changed from fast fitting to stable convergence. At the same time, the accuracy rate of verification continues to improve with the progress of training, and the growth rate stabilizes after the 20th round, and stabilizes at about 89.7% around the 100th round, without obvious decline, indicating that the model does not produce obvious overfitting while achieving high recognition ability.

From the overall trend, the three curves of training loss, validation loss and verification accuracy maintain a good correspondence: the rate of decline and increase is fast in the early stage, the change is gradual in the middle and late stage, and there is only a small fluctuation in the later stage, which indicates that the model in this paper has good robustness and system stability in the process of parameter updating and feature learning. In summary, the model achieves stable convergence after 100 rounds of training, in which the training loss reduction reaches 83.9%, the validation loss reduction reaches 79.8%, and the validation accuracy finally reaches 89.7%, indicating that the method can maintain good generalization performance while ensuring training efficiency, which provides reliable support for subsequent teaching applications.

5.5 Analysis on the application effect of English Writing teaching in Higher Vocational Colleges

In order to further verify the practical application value of this method in the English writing class of higher vocational colleges, this paper compared the writing performance of the experimental group and the control group before and after the teaching implementation, and comprehensively evaluated the students 'acceptance of the system feedback and the change of revision gain. The relevant results are shown in Table 4.

Table 4: Comparison of writing performance between experimental group and control group

Group	Test Stage	Language Accuracy	Semantic Expression	Discourse Coherence	Task Completion	Total Score
Control Group	Pre-test	71.6	69.8	68.9	70.4	70.2
Control Group	Post-test	75.3	73.1	72.4	74.0	73.7
Experimental Group	Pre-test	71.9	70.2	69.1	70.6	70.5
Experimental Group	Post-test	82.8	81.4	80.3	82.1	81.7

As can be seen from Table 4, the total scores of the two groups of students in the pre-test were close, with 70.2 and 70.5 in the control group and the experimental group, respectively, showing a small basic difference. After a period of teaching, the total score of the post-test in the control group increased to 73.7, with an increase of 3.5. In the experimental group, the total score of the post-test reached 81.7, which was 11.2 points higher than that of the pre-test, and the improvement was significantly greater. Among them, the experimental group improved 10.9 points, 11.2 points, 11.2 points and 11.5 points respectively in the four indicators of language accuracy, semantic expression, discourse coherence and task completion, indicating that the feedback mechanism based on deep semantic understanding not only improved the recognition of local language problems, but also had a strong effect on the overall expression organization and task achievement of students.

Student feedback adoption rate versus revision gain variation is shown in Figure 5.

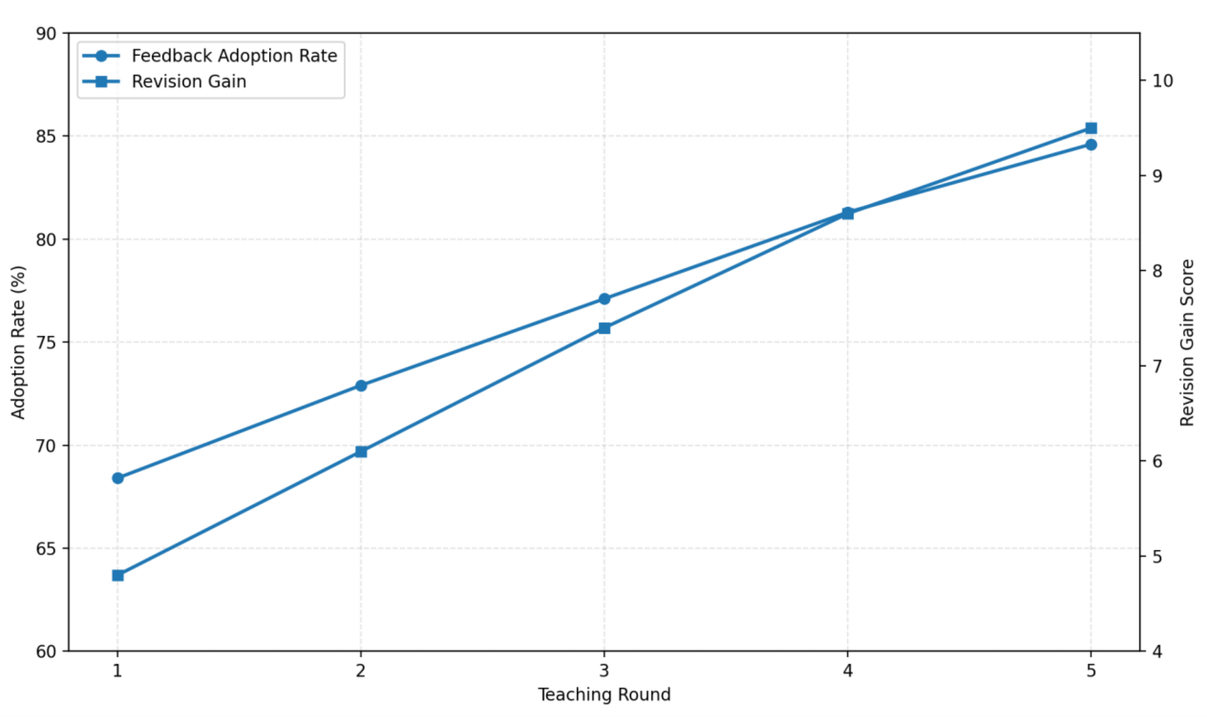

Figure 5: Student feedback adoption rate and revision gain change diagram

As shown in Figure 5, with the advancement of teaching rounds, the adoption rate and revision gain of students 'system feedback show a continuous upward trend. In the first round, the feedback acceptance rate was 68.4%, and the revision gain score was 4.8. In the fifth round, the two indexes increased to 84.6% and 9.5 respectively. In general, the feedback adoption rate increased by 16.2 percentage points, and the revision gain increased by 4.7 points, indicating that students 'understanding and utilization ability of intelligent feedback was continuously enhanced, and the system feedback had gradually transformed from the initial auxiliary hint to an important support for promoting writing improvement.

6 Conclusion

Aiming at the problems of feedback lag, insufficient semantic diagnosis and weak teaching adaptability in English writing in higher vocational colleges, this paper proposes an AI English writing feedback mechanism based on deep semantic understanding, and completes the overall construction from corpus preprocessing, deep feature construction, feedback task collaborative modeling to teaching implementation process design. The results show that the joint modeling of lexical, syntactic, semantic and discourse features can more completely support writing evaluation, and the semantic features and discourse features have a more obvious effect on improving the comprehensive feedback quality. Multi-task optimization makes the system form a more stable cooperative relationship among bias identification, semantic interpretation and feedback generation. The experimental results show that the writing evaluation accuracy of the complete model reaches 89.7%, the comprehensive feedback quality score reaches 88.6, and the verification accuracy is stable at about 89.7% after training, which reflects good convergence and generalization ability. In the teaching application level, the total score of the experimental group was 11.2 points higher than that of the pre-test, indicating that this method can effectively transform technical feedback into writing improvement effect. In the future, it can continue to combine richer corpus of career scenes and personalized recommendation mechanism to further improve the transfer ability and classroom adaptation level of the system.

References

[1] Li Z. Teachers in automated writing evaluation (AWE) system-supported ESL writing classes: Perception, implementation, and influence[J]. System, 2021, 99: 102505. DOI: 10.1016/j.system.2021.102505

[2] Ranalli J. L2 student engagement with automated feedback on writing: Potential for learning and issues of trust[J]. Journal of Second Language Writing, 2021, 52: 100816. DOI: 10.1016/j.jslw.2021.100816

[3] Hassanzadeh M, Fotoohnejad S. Implementing an automated feedback program for a foreign language writing course: A learner-centric study[J]. Journal of Computer Assisted Learning, 2021, 37(5): 1494-1507. DOI: 10.1111/jcal.12587

[4] Nunes A, Cordeiro C, Limpo T, Castro S L. Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020[J]. Journal of Computer Assisted Learning, 2022, 38(2): 599-620. DOI: 10.1111/jcal.12635

[5] Guo Q, Feng R, Hua Y. How effectively can EFL students use automated written corrective feedback (AWCF) in research writing?[J]. Computer Assisted Language Learning, 2022, 35(9): 2312-2331. DOI: 10.1080/09588221.2021.1879161

[6] Yu S, Zhang E D, Liu C. Assessing L2 student writing feedback literacy: A scale development and validation study[J]. Assessing Writing, 2022, 53: 100643. DOI: 10.1016/j.asw.2022.100643

[7] Li R. Still a fallible tool? Revisiting effects of automated writing evaluation from activity theory perspective[J]. British Journal of Educational Technology, 2023, 54(3): 773-789. DOI: 10.1111/bjet.13294

[8] Shi H, Aryadoust V. A systematic review of AI-based automated written feedback research[J]. ReCALL, 2024, 36(2): 187-209. DOI: 10.1017/S0958344023000265

[9] Karatay Y, Karatay L. Automated writing evaluation use in second language classrooms: A research synthesis[J]. System, 2024, 123: 103332. DOI: 10.1016/j.system.2024. 103332

[10] Sari E, Han T. The impact of automated writing evaluation on English as a foreign language learners' writing self-efficacy, self-regulation, anxiety, and performance[J]. Journal of Computer Assisted Learning, 2024, 40(5): 2065-2080. DOI: 10.1111/jcal. 13004

[11] Fu Q K, Zou D, Xie H, Cheng G. A review of AWE feedback: types, learning outcomes, and implications[J]. Computer Assisted Language Learning, 2024, 37(1-2): 179-221. DOI: 10.1080/09588221.2022.2033787

[12] Saricaoglu A, Bilki Z. Voluntary use of automated writing evaluation by content course students[J]. ReCALL, 2021, 33(3): 265-277. DOI: 10.1017/S0958344021000021

[13] Reynolds B L, Kao C W, Huang Y Y. Investigating the effects of perceived feedback source on second language writing performance: A quasi-experimental study[J]. The Asia-Pacific Education Researcher, 2021, 30(6): 585-595. DOI: 10.1007/s40299-021-00597-3

[14] Wang Z, Han F. The effects of teacher feedback and automated feedback on cognitive and psychological aspects of foreign language writing: A mixed-methods research[J]. Frontiers in Psychology, 2022, 13: 909802. DOI: 10.3389/fpsyg.2022.909802

[15] Wei P, Wang X, Dong H. The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial[J]. Frontiers in Psychology, 2023, 14: 1249991. DOI: 10.3389/fpsyg.2023.1249991

[16] Ajabshir Z F, Ebadi S. The effects of automatic writing evaluation and teacher-focused feedback on CALF measures and overall quality of L2 writing across different genres[J]. Asian-Pacific Journal of Second and Foreign Language Education, 2023, 8(1): 26. DOI: 10.1186/s40862-023-00201-9

[17] Barrot J S. Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy[J]. Computer Assisted Language Learning, 2023, 36(4): 584-607. DOI: 10.1080/09588221.2021.1936071

[18] Fan N. Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study[J]. SAGE Open, 2023, 13(2): 1-17. DOI:

10.1177/21582440231181296

[19] Dizon G, Gayed J M. A systematic review of Grammarly in L2 English writing contexts[J]. Cogent Education, 2024, 11(1): 2397882. DOI: 10.1080/2331186X.2024. 2397882

[20] Zhang Z, Hyland K. The role of digital literacy in student engagement with automated writing evaluation (AWE) feedback on second language writing[J]. Computer Assisted Language Learning, 2025, 38(5-6): 1060-1085. DOI: 10.1080/09588221.2023.2256815

[21] Huang Y, Wilson J. Exploring the effectiveness of large-scale automated writing evaluation implementation on state test performance using generalised boosted modelling[J]. Journal of Computer Assisted Learning, 2025, 41(2): e70009. DOI: 10.1111/jcal.70009

[22] Jiang Y, Beigman Klebanov B, Hao J, Deane P, Livne O E. Unveiling patterns of interaction with automated feedback in Writing Mentor and their relationships with use goals and writing outcomes[J]. Journal of Computer Assisted Learning, 2025, 41(2): 1-15. DOI: 10.1111/jcal.70014

[23] Muñoz Muñoz B C, Nassaji H, Bello Carrillo F I. ChatGPT-generated versus human direct corrective feedback on L2 writing[J]. System, 2025, 134: 103805. DOI: 10.1016/j.system.2025.103805

[24] Li X, Ke P. How L2 student writers engage with automated feedback: A longitudinal perspective[J]. Assessing Writing, 2025, 64: 100919. DOI: 10.1016/j.asw.2025.100919

[25] Rahimi M, Fathi J, Zou D. Exploring the impact of automated written corrective feedback on the academic writing skills of EFL learners: An activity theory perspective[J]. Education and Information Technologies, 2025, 30(3): 2691-2735. DOI: 10.1007/s10639-024-12896-5

[26] Ekizoğlu M, Demir A N. The role of AI assisted writing feedback in developing secondary students writing skills[J]. Discover Education, 2025, 4(1): 454. DOI: 10.1007/s44217-025-00919-3