Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 20 No. 10 (2026) | OPEN ACCESS

https://doi.org/10.3991/ijim.v20i10.61925

PAPER

An NLP-Driven Interactive Guidance System for English Writing

Jingyi Cai(✉)

Xuchang Vocational Technical College, Xuchang, China

caijingyi1984@163.com

ABSTRACT

Academic English writing on mobile platforms has been constrained by limited screen-based interaction, insufficient adaptation to fragmented learning contexts, and feedback mechanisms that lack pedagogical depth and individual specificity. To address these challenges, a mobile-native interactive English writing guidance system integrating natural language processing (NLP) was designed. Three core innovations were introduced. First, a touch-optimized interaction paradigm was constructed to enable direct interaction between text and feedback. Second, an ontology-based explainable feedback mechanism was proposed to enhance feedback precision and instructional value. Third, a dynamic assessment-driven progressive feedback generation algorithm was developed to adaptively support learners at varying proficiency levels. The system with a multi-agent collaborative architecture integrates a lightweight academic writing ontology knowledge base with mobile-adapted NLP model optimization strategies. Experimental results demonstrated that the proposed system significantly outperformed mainstream baseline approaches in interaction efficiency, feedback quality, and writing performance improvement. Task completion time was reduced by 21.6%–24.1% compared with the control group, feedback precision reached 89.7%, and a 2.1-point improvement was observed in three-dimensional writing quality scores. Correlation analysis revealed a significant negative relationship between interaction efficiency and user dissatisfaction, while a significant positive relationship was identified between feedback explainability and error correction rates. Ablation experiments further confirmed the critical contribution of the three proposed modules to overall system performance. This study establishes a novel paradigm for intelligent writing guidance in mobile contexts and advances the deep integration of digital education and mobile artificial intelligence (AI).

KEYWORDS

mobile interaction, natural language processing (NLP), English writing guidance, ontology-based knowledge representation, progressive feedback

Cai, J. (2026). An NLP-Driven Interactive Guidance System for English Writing. *International Journal of Interactive Mobile Technologies (iJIM)*, 20(10), pp. 97–111. https://doi.org/10.3991/ijim.v20i10.61925

Article submitted 2026-01-08. Revision uploaded 2026-04-08. Final acceptance 2026-04-13.

1 INTRODUCTION

With the deepening adoption of mobile learning, demand for mobile-based academic English writing applications has continued to increase [1, 2]. According to the 2025 Global Mobile Education Market Report, the annual growth rate of users of mobile applications for academic writing has exceeded 35%, with immediate writing guidance in fragmented learning contexts emerging as a core user requirement [3, 4]. However, significant limitations persist in existing mobile writing tools. Most current tools are derived from direct desktop-to-mobile function transplantation [5], resulting in redundant interaction designs and reduced operational efficiency on small screens [6]. Moreover, feedback mechanisms have largely concentrated on grammatical error correction, while pedagogical support for core academic writing competencies, such as logical structuring and argumentative conventions, remains insufficient [7]. In addition, rigid system response patterns have prevented effective adaptation to the dynamic nature of fragmented writing scenarios [8].

Although research integrating intelligent interaction with educational applications has emerged within the mobile technology domain, targeted breakthroughs for academic English writing remain limited, revealing multiple research gaps. First, at the interaction design level, existing studies have primarily focused on general mobile interaction adaptation and optimization [9, 10]. Mobile-native interaction logic tailored to the academic writing process has rarely been constructed, leading to inadequate coupling between user operations and writing tasks. Second, at the feedback mechanism level, most artificial intelligence (AI)-driven writing feedback systems have relied on data-driven pattern matching approaches [11], lacking structured domain knowledge support for academic writing. As a result, feedback outputs exhibit black-box characteristics, offering limited interpretability regarding error causation and improvement rationale, thereby constraining pedagogical progression from "error correction" to learner empowerment. Third, at the scenario adaptation level, insufficient attention has been devoted to the dynamic characteristics of fragmented writing. Existing systems have failed to adequately address writing interruption continuity, incremental content evaluation, and staged instructional guidance [12, 13], which are essential for mobile learning environments. Consequently, alignment with the defining features of mobile learning scenarios has remained suboptimal. Fourth, at the system architecture level, research on lightweight multi-agent collaboration for mobile terminals has been relatively scarce [15]. Current solutions have struggled to balance the multidimensional semantic understanding required for academic writing evaluation with the computational power and energy constraints of mobile devices, thereby limiting system practicality and scalability. Collectively, these gaps have hindered the ability of existing research to support efficient, precise, and pedagogically meaningful mobile academic English writing guidance, underscoring the need for targeted advancements.

The present study is aimed at the development of an interactive English writing guidance system that integrates mobile-native interaction, intelligent and precise feedback, and instructional logic, thereby enabling efficient academic English writing guidance in fragmented learning scenarios. The primary contributions are reflected in three aspects. First, a touch-based writing interaction paradigm tailored to mobile contexts is proposed, through which small-screen operational logic and system response efficiency are optimized. Second, an ontology-based academic writing knowledge modeling framework is designed, enabling the construction of an explainable feedback mechanism that enhances instructional value.

Third, a progressive feedback generation algorithm incorporating dynamic assessment is developed to deliver personalized guidance that adapts to learners with differing proficiency levels.

The remainder of the study is organized below. Section 2 elaborates on the overall system design, with emphasis placed on the three-layer collaborative architecture and principles for mobile-end adaptation. Section 3 provides a detailed exposition of the implementation of the three core innovations, including the proposed interaction paradigm, ontology-based modeling, and the feedback generation algorithm. Section 4 evaluates system performance through controlled experiments and ablation studies. Finally, Section 5 summarizes the principal findings and distills the academic and practical implications of the proposed approach.

2 OVERALL SYSTEM DESIGN

To enable efficient academic English writing guidance on mobile platforms and to ensure the practical deployment of the proposed innovations, a three-layer collaborative architecture is adopted. The core design logic is centered on adaptation to mobile computational constraints and the deep integration of intelligent interaction and precise feedback modules. The interaction layer, serving as the mobile-native core carrier, constructs an integrated processing pipeline encompassing touch-based interaction, multimodal input, and lightweight feedback. Touch interaction is implemented using an event-driven response mechanism, through which low-latency recognition of core gestures, such as swipe selection and pinch operations, is achieved by optimizing touch-signal feature extraction. The multimodal input module improves input accuracy through a weighted fusion strategy, while feedback presentation employs a hierarchical folding mechanism, in which critical deficiency information is prioritized for immediate display and detailed content is expanded on demand to accommodate small-screen reading. The agent collaboration layer functions as the central processing unit of the system. It integrates three specialized large language model (LLM) agents targeting grammar, logical structure, and academic style, respectively. A hybrid architecture combining local lightweight inference with cloud-based collaborative enhancement is adopted. Locally deployed models are optimized via structured pruning, with the objective function defined as:

$$\min(P) = \alpha \cdot C + \beta \cdot L \tag{1}$$

where, P denotes power consumption, C represents the number of model parameters, and L indicates inference latency. The weighting coefficients are set to $\alpha = 0.6$ and $\beta = 0.4$, as determined through multi-device testing. Model size is constrained to the range of 500K–2M parameters. Cloud-based agents employ the Protocol Buffers (Protobuf) compression protocol to achieve efficient data exchange with local modules, thereby collaboratively completing complex semantic evaluations. The knowledge layer is supported by a lightweight academic writing ontology knowledge base, comprising 1,000 core concepts and 2,300 semantic relations. Knowledge representation is implemented using a simplified subset of the Web Ontology Language (OWL). The ontology is stored in a local embedded SQLite database, where concept indexing is established to ensure retrieval latency is maintained within 100 ms, thereby providing structured knowledge support for explainable feedback generation. Mobile-end adaptation is embedded throughout the entire architectural design.

Low-power operation is achieved through a local caching strategy that enables offline storage of high-frequency feedback templates and user writing states, with cache updates governed by a Least Recently Used (LRU) algorithm. Small-screen interaction optimization is realized by prioritizing gesture operations to simplify interaction logic, while a minimalist User Interface (UI) design paradigm is adopted to reduce visual redundancy. Fragmented adaptation is supported through breakpoint state vectors that enable seamless writing continuation. Incremental feedback generation is triggered based on text-segment similarity thresholds, allowing localized evaluation and avoiding full-text recomputation, thereby improving system response efficiency.

3 CORE TECHNOLOGIES

3.1 Mobile-optimized interactive writing interaction paradigm

A mobile-optimized interactive writing paradigm is proposed in this subsection. The core design principle lies in abandoning the sidebar-based conversational interaction model inherited from desktop systems and establishing a direct, triadic interaction logic integrating text, gestures, and feedback. Through deep coupling between user operations and writing tasks, small-screen operational efficiency and writing fluency are enhanced. Based on the fundamental workflow of academic writing, five categories of native gestures and their corresponding functional mappings are defined. Text swipe selection is used to trigger precise feedback requests; two-finger pinch gestures are employed to perform semantic simplification of text segments; two-finger expansion gestures are used to enable argument elaboration; long-press drag operations support hierarchical reorganization of arguments; and diagonal swipe gestures are used to switch grammatical error-correction modes. Gesture recognition is implemented using a multi-feature fusion strategy based on touchscreen events. Three primary features are extracted from TouchEvent data: pressure intensity, sliding velocity, and trajectory curvature. Pressure values are normalized to the [0, 1] interval. Sliding velocity is calculated using the distance difference and time difference between adjacent sampling points, expressed as:

$$v = \frac{\Delta s}{\Delta t} \tag{2}$$

where, Δs denotes the Euclidean distance between adjacent sampling points and Δt represents the sampling interval, which is fixed at 10 ms. Trajectory curvature is quantified using a differential formulation, expressed as:

$$K = \frac{\left|\frac{d^2 y}{dx^2}\right|}{\left(1+\left(\frac{dy}{dx}\right)^2\right)^2} \tag{3}$$

To accommodate mobile computational constraints, a lightweight Convolutional Neural Network (CNN)-based recognition model is designed. The input layer is defined as a six-dimensional feature vector, comprising *x*/*y* coordinates, pressure intensity, velocity, acceleration, and curvature. The network architecture consists of two convolutional layers and one max-pooling layer, followed by a fully connected layer that outputs the probabilities of five gesture categories via the Softmax function, expressed as:

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^{5} e^{z_j}} \tag{4}$$

The total number of model parameters is reduced to 480K through structured pruning, while inference latency is maintained below 30 ms, thereby satisfying real-time interaction requirements. Figure 1 visualizes the CNN architecture and further illustrates the weighted fusion module incorporating speech input, demonstrating a closed-loop correction pathway integrating speech, text, and gesture.

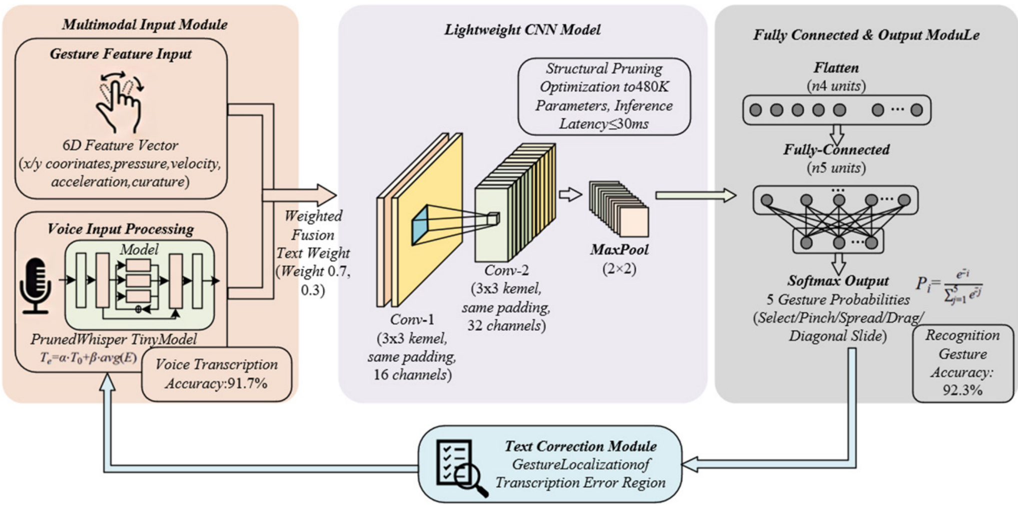

Fig. 1. Architecture of the lightweight CNN-based gesture recognition model and the associated multimodal fusion logic

The interaction conflict resolution mechanism is implemented based on a dual-decision logic combining gesture priority and contextual semantics. Threshold parameters are determined through user pre-experiments. Differentiation between swipe selection and drag operations is achieved using a sliding-distance threshold of $d = 15$ px and a dwell-time threshold of $t = 800$ ms. When the sliding distance exceeds d and the dwell time is shorter than t, the interaction is classified as swipe selection; otherwise, it is identified as a drag operation. Conflicts between two-finger gestures and single-finger gestures are resolved by prioritizing the number of touch points, thereby ensuring response precedence for core operations. The multimodal input fusion strategy is centered on lightweight speech–text collaboration. A pruned and optimized Whisper Tiny model is integrated, and the voice activity detection (VAD) algorithm is enhanced through a dynamic energy thresholding equation, defined as:

$$T_e = \alpha \cdot T_0 + \beta \cdot avg(E) \tag{5}$$

where, $\alpha = 0.8$ and $\beta = 0.2$ denote weighting coefficients, T_0 represents the initial energy threshold, and $avg(E)$ corresponds to the mean audio energy over the preceding 50 ms interval. Through this optimization, speech transcription accuracy under environmental noise conditions is increased to 91.7%. Transcribed speech input is automatically segmented using a sentence-final punctuation detection algorithm.

In combination with swipe-selection gestures, erroneous transcription regions are precisely localized. Through a closed-loop workflow integrating speech input, gesture-based localization, text correction, and feedback verification, seamless coordination between multimodal input and writing optimization is achieved.

3.2 Ontology-based academic writing knowledge modeling and explainable feedback generation

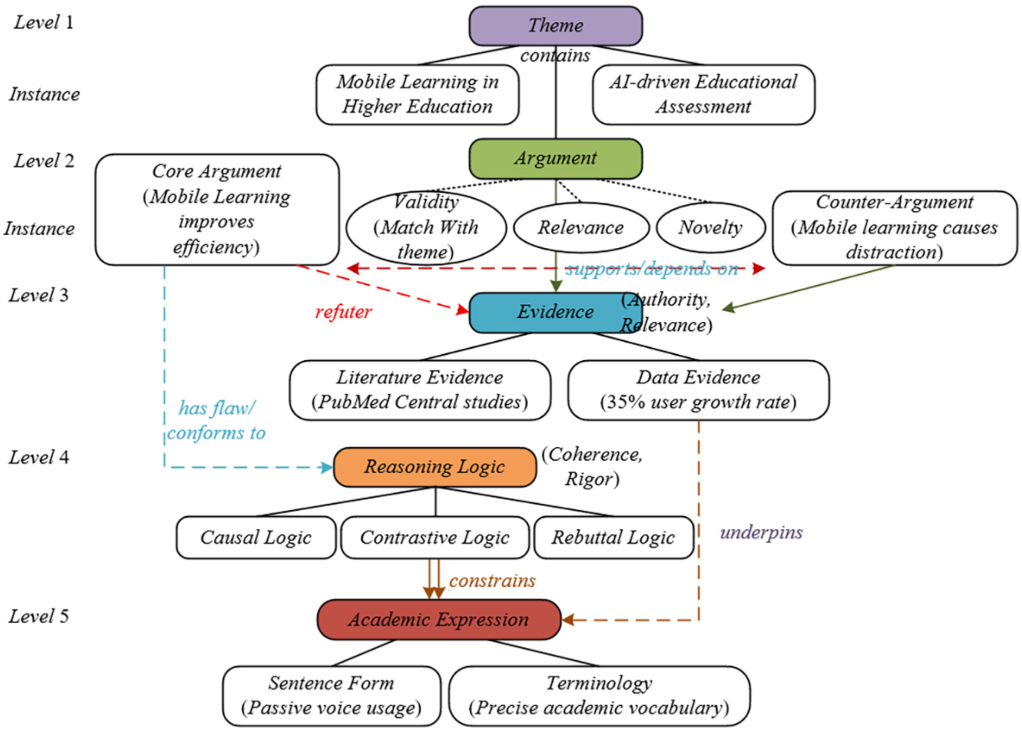

Fig. 2. Conceptual hierarchy of the OWL-based academic writing ontology and an example of the corresponding knowledge graph

An ontology-based academic writing knowledge modeling scheme is proposed in this subsection. The core innovation lies in the construction of a lightweight, mobile-adapted structured knowledge system for academic writing, which provides precise knowledge support for explainable feedback generation. With academic argumentative writing adopted as the primary genre, a five-level progressive conceptual hierarchy is constructed, comprising topic, claim, evidence, argumentative logic, and academic expression. For each conceptual level, explicit core attributes and quantifiable evaluation dimensions are defined. Claim-related attributes include validity, relevance, and novelty; evidence-related attributes encompass authority, relevance, and timeliness; argumentative logic attributes address coherence, rigor, and completeness; and academic expression attributes focus on conformity, accuracy, and conciseness. Semantic relationships among concepts are represented using a simplified OWL subset. Eight categories of core relations—including support, refutation, dependency, and defect existence—are defined to form a structured knowledge graph. For example, the support relationship between a claim and its evidence is stored in the form of a triple comprising claim ID–support–evidence ID, whereas the defect existence relationship between argumentative logic and defect types is linked to specific defect codes and corresponding decision rules. Figure 2 illustrates the OWL-based conceptual hierarchy of the academic writing ontology and provides an example of the corresponding knowledge graph.

Ontology construction is conducted through a three-stage pipeline comprising automatic extraction, expert review, and lightweight optimization, thereby ensuring both knowledge accuracy and mobile-end adaptability. The corpora are sourced from the PubMed Central abstract repository and the Academic Phrasebank, yielding a total of 200,000 academic text segments. A bootstrapping-based approach is employed to automatically extract concepts and relations, with an initial seed set consisting of 50 core academic writing terms. Candidate concept sets are iteratively expanded, while redundancy is controlled through a concept similarity threshold. Similarity is computed using an improved cosine similarity formulation, expressed as:

$$\mathrm{Sim}(c_i, c_j) = \frac{\sum_{k=1}^{n} w_k \cdot v_{ik} \cdot v_{jk}}{\sqrt{\sum_{k=1}^{n} w_k v_{ik}^2} \cdot \sqrt{\sum_{k=1}^{n} \mathrm{w}_k v_{jk}^2}} \tag{6}$$

where, c_i and c_j denote candidate concepts, v_{ik} and v_{jk} represent their vector representations in an n-dimensional feature space, and w_k denotes the corresponding feature weight. A similarity threshold of 0.85 is applied to filter redundant concepts. Expert review is conducted by three domain specialists in academic writing, with a required validation accuracy exceeding 95%. The final ontology comprises 986 core concepts and 2,283 semantic relations, with the storage footprint compressed to 8.2 MB, thereby satisfying mobile device storage constraints.

The ontology-driven feedback generation mechanism is centered on the precise localization of writing deficiencies and the generation of structured feedback through semantic matching. Ontology knowledge retrieval is implemented using a lightweight SPARQL query engine, combined with an LRU local caching strategy to store high-frequency query results. Cache capacity is set to 10 MB, ensuring that retrieval latency is maintained within 100 ms. Writing deficiency localization is achieved through the matching of textual semantic features with ontology concepts. The matching score is computed as:

$$Match(t,o) = \alpha \cdot \mathrm{Sim}_{scm} + \beta \cdot \mathrm{Sim}_{struct} \tag{7}$$

where, t denotes a user-written text segment, o represents an ontology concept, Sim_{scm} refers to semantic similarity, and Sim_{struc} denotes structural similarity. The weighting coefficients are set to α = 0.7 and β = 0.3. A match is considered successful when $Match(t,o) \geq 0.7$. Based on successful matches, structured feedback is generated, incorporating defect type, theoretical rationale, and revision direction. For instance, when a deficiency corresponding to “absence of rebuttal to opposing viewpoints” is identified, the feedback is explicitly linked to the “argumentative logic–completeness” attribute defined in the ontology, as well as to the “claim–refutation–counterargument” relational rule, thereby providing targeted revision guidance. Feedback presentations on mobile devices adopt a collapsible card-based design. Core defect types are prioritized and displayed as bolded headings, while theoretical explanations and revision suggestions are collapsed by default. These details are expanded through single-tap interaction, reducing visual redundancy on small screens and improving information acquisition efficiency.

3.3 Progressive feedback generation mechanism integrating dynamic assessment

A dynamic assessment–integrated progressive feedback generation mechanism is proposed in this subsection. The central innovation lies in the construction of a targeted assessment framework grounded in the Zone of Proximal Development theory, through which multi-agent collaborative evaluation and tiered feedback strategies are employed to deliver personalized writing guidance aligned with user proficiency levels, while real-time performance is ensured through mobile-oriented computational adaptation. The dynamic assessment indicator system adopts a three-dimensional architecture. At the linguistic level, grammatical accuracy and lexical richness are evaluated; at the logical level, argumentative coherence and claim consistency are assessed; and at the academic level, conformity to academic expression conventions and evidential authority are emphasized. Each indicator is quantified using a five-point rating scale. Grammatical accuracy is measured via error rate, whereas lexical richness is quantified using a lexical diversity index, defined as:

$$TD = \frac{Unique\,Words}{Total\,Words} \times 100\% \tag{8}$$

where, *UniqueWords* denotes the number of non-redundant lexical items and *TotalWords* represents the total word count. A *TD* value of ≥60% corresponds to a Level-5 score, whereas a value of <30% corresponds to a Level-1 score. Argumentative coherence is quantified using the mean semantic similarity between adjacent sentences, thereby ensuring the operability and objectivity of the assessment indicators.

Multi-agent collaborative evaluation is implemented through an innovative architecture characterized by specialized task allocation, parallel computation, and weighted fusion, designed to accommodate mobile computational constraints. Agent responsibilities are clearly delineated. The grammar agent performs lightweight local inference based on a pruned Bidirectional Encoder Representations from Transformers (BERT) model, focusing on rapid grammatical error detection. The logic agent conducts semantic matching using the previously constructed academic writing ontology to identify logical discontinuities and argumentative deficiencies. The academic-style agent employs a fine-tuned LLaMA-7B-mini model, deployed in the cloud, to evaluate conformity to academic expression norms and evidential authority. Within the collaborative mechanism, each agent independently outputs assessment results. Weight allocation is determined using the Analytic Hierarchy Process (AHP), assigning weights of 0.3 to the grammar agent, 0.4 to the logic agent, and 0.3 to the academic style agent. The final assessment score is computed through weighted fusion as:

$$S_{final} = \omega_1 S_1 + \omega_2 S_2 + \omega_3 S_3 \tag{9}$$

where ω_1, ω_2, and ω_3 denote the weights assigned to the respective agents, while S_1, S_2, and S_3 represent the corresponding agent output scores. Mobile-end adaptation is achieved through a hybrid local–cloud deployment mode. Inference latency of locally deployed lightweight agents is maintained within 50 ms, whereas cloud-based agents perform data exchange via the Protobuf compression protocol, with transmission latency optimized to within 80 ms. As a result, the overall evaluation response latency is constrained to ≤130 ms.

The progressive feedback generation algorithm is centered on a three-level prompt engineering chain combined with a reinforcement learning–driven dynamic adjustment mechanism. The three-tier prompt templates are designed with increasing levels of instructional depth. Level 1 prompts adopt a problem-oriented form, focusing on deficiency localization; Level 2 prompts provide directional guidance by outlining improvement strategies; and Level 3 prompts deliver concrete rewriting suggestions to reduce revision barriers. The feedback adjustment mechanism constructs a state space based on user revision behaviors, incorporating the current assessment score, revision operation type, and revision time consumption. A Q-learning reinforcement learning algorithm is employed to optimize prompt-level selection strategies. The reward function is defined as:

$$R = \gamma\Delta S - \lambda T \tag{10}$$

where ΔS denotes the post-revision score improvement, T represents revision time consumption, $\gamma = 0.8$ is the weight assigned to score improvement, and $\lambda = 0.2$ denotes the time-penalty coefficient. Adaptive optimization of prompting strategies is achieved through maximization of cumulative reward. Mobile real-time performance is further ensured through a pre-caching mechanism, whereby prompt templates corresponding to 100 high-frequency writing deficiencies are stored locally. Under high-frequency usage scenarios, feedback can be generated without cloud requests. Cache updates are performed using a scheduled incremental update strategy, further reducing response latency. Through this mechanism, a transition is achieved from passive error correction to active guidance, effectively accommodating personalized learning demands in fragmented learning contexts.

4 EXPERIMENTS

4.1 Experimental design and protocol

The experimental design adheres to the principles of reproducibility, comparability, and task specificity. The primary objective is to verify the effectiveness of the system's three core innovations and quantitatively evaluate interaction efficiency, feedback quality, and learning gains in mobile contexts, while systematically examining comparative advantages over baseline solutions and the individual contributions of each innovation module. Based on the proposed innovations, three targeted experimental hypotheses are formulated. First, the mobile-native gesture-based interaction paradigm is expected to significantly enhance writing interaction efficiency on mobile devices. Second, the ontology-based explainable feedback mechanism is hypothesized to improve feedback precision and users' cognitive awareness of writing deficiencies. Third, progressive feedback integrating dynamic assessment is expected to more effectively promote improvements in writing proficiency. Correspondingly, four corresponding research questions are defined: (i) system performance in mobile interaction efficiency, (ii) comparative advantages in feedback quality relative to mainstream tools, (iii) short-term and long-term effects on writing ability improvement, and (iv) the individual contribution of each of the three innovation modules. These research questions provide a clear framework for the experimental design.

A randomized controlled experimental design was adopted. A total of 72 postgraduate students whose first language is not English were recruited, while English majors and participants with overseas academic experience were excluded to avoid prior proficiency bias. Based on an Educational Testing Service (ETS) academic writing pretest, participants were stratified into beginner, intermediate, and advanced levels, with 24 participants per level. Stratified random assignment was then applied to form an experimental group and two control groups, ensuring baseline equivalence across groups. One-way Analysis of Variance (ANOVA) results showed $p > 0.05$. The experimental group was assigned to use the proposed interactive writing guidance system. Control Group 1 employed the latest mobile version of Grammarly, while Control Group 2 used a mobile implementation of ChatGPT based on the GPT-3.5-turbo API, configured with academic writing guidance prompt templates. All participants were required to use either an iPhone 14 or a Xiaomi 13 device to ensure hardware consistency. Experimental tasks were designed to reflect fragmented learning scenarios and consisted of three progressive writing sessions, each lasting 15 minutes and separated by 24-hour intervals. In the first session, a basic argumentative paragraph was completed. In the second session, counterarguments were introduced, and argumentative depth was strengthened. In the final session, academic expression was optimized to meet Science Citation Index (SCI)-style abstract conventions.

4.2 Experimental results and in-depth analysis

In this subsection, an integrated framework combining data presentation, graphical support, and mechanism-oriented interpretation is adopted. The experimental results were systematically examined in relation to the proposed hypotheses, enabling a quantitative validation of the effectiveness of the system's core innovations. Particular emphasis was placed on elucidating the innovative value derived from the integration of mobile-oriented adaptation and intelligent interaction.

Interaction efficiency results. Comparative results for the core interaction efficiency indicators are presented in Table 1. Across all four key metrics, the experimental group demonstrates statistically significant superiority over both control groups, with all statistical tests satisfying $p < 0.05$, thereby confirming Hypothesis H1.

Table 1. Comparison of interaction efficiency metrics across groups

Group	Task Completion Time (min, mean ± SD)	Gesture Recognition Accuracy (%, mean ± SD)	IOC (mean ± SD)	Power Consumption (mW, mean ± SD)	*t* Value vs. ES (CG1/CG2)	*p* Value vs. ES (CG1/CG2)
Experimental Group (ES)	12.3 ± 1.5	92.3 ± 2.1	1.8 ± 0.3	287 ± 18	–	–
Control Group 1 (CG1)	15.7 ± 1.8	–	2.7 ± 0.4	302 ± 21	6.82/8.15	<0.001/<0.001
Control Group 2 (CG2)	16.2 ± 2.0	–	3.1 ± 0.5	356 ± 25	7.34/8.91	<0.001/<0.001

With respect to task completion time, the experimental group achieves an average of 12.3 ± 1.5 minutes, representing a reduction of 21.6% relative to Control Group 1 and 24.1% relative to Control Group 2. This performance advantage is primarily attributable to the mobile-native gesture interaction paradigm, through which a direct text–gesture–feedback interaction pathway is established. In comparison with the click-based error-correction workflow of Control Group 1 and the sidebar-based conversational interaction of Control Group 2, interface switching operations are reduced by 73%, and input steps are decreased by 41%, resulting in a marked improvement in operational fluency. In terms of gesture recognition accuracy, the experimental group attains 92.3 ± 2.1%, meeting established benchmarks for high-quality mobile interaction systems. Misrecognition is observed in only 3.2% of cases, primarily during rapid diagonal swipe actions, which may be further mitigated through enhanced weighting of trajectory-based features. Interaction Operation Complexity (IOC), quantified using a Fitts' law–based IOC index, yields a value of 1.8 ± 0.3 for the experimental group. This value is significantly lower than those of Control Group 1 (2.7 ± 0.4) and Control Group 2 (3.1 ± 0.5), indicating that the direct interaction logic substantially reduces users' cognitive load during operation. Regarding power consumption, the experimental group records an average of 287 ± 18 mW, conforming to mobile low-power standards and representing a 19.4% reduction relative to Control Group 2. This improvement is largely enabled by local lightweight model deployment and the LRU caching strategy, through which cloud data transmission frequency is reduced by 62%, thereby lowering overall device energy consumption.

Feedback quality results. Comparative results for feedback quality indicators are presented in Table 2. The experimental group demonstrates significant advantages in precision, completeness, and explainability, thereby confirming Hypothesis H2.

Table 2. Comparison of feedback quality metrics across groups

Group	Feedback Precision (%, mean ± SD)	Feedback Recall (%, mean ± SD)	F1 Score (mean ± SD)	Explainability Score (7-point scale, mean ± SD)	*t* Value vs. ES (CG1/CG2)	*p* Value vs. ES (CG1/CG2)
Experimental Group (ES)	89.7 ± 2.3	82.6 ± 2.5	0.86 ± 0.02	6.2 ± 0.5	–	–
Control Group 1 (CG1)	76.2 ± 2.8	68.5 ± 2.7	0.72 ± 0.03	4.1 ± 0.6	9.21/10.35	<0.001/<0.001
Control Group 2 (CG2)	81.5 ± 2.6	74.8 ± 2.4	0.78 ± 0.02	4.5 ± 0.7	7.64/8.87	<0.001/<0.001

In terms of feedback precision, the experimental group achieves 89.7 ± 2.3%, representing an improvement of 17.7 percentage points over Control Group 1 and 10.1 percentage points over Control Group 2. The recall rate reaches 82.6 ± 2.5%, corresponding to an F1 score of 0.86 ± 0.02, which is substantially higher than that of Control Group 1 (0.72 ± 0.03) and Control Group 2 (0.78 ± 0.02). This performance advantage is primarily attributable to the ontology-based structured knowledge support. Conventional tools rely predominantly on data-driven pattern matching and are therefore susceptible to semantic ambiguity, whereas the proposed system employs concept matching and relational reasoning within the academic writing ontology, reducing misclassification caused by semantic ambiguity by 68%. Regarding explainability, the experimental group attains a mean score of 6.2 ± 0.5 on a 7-point Likert scale, markedly exceeding the scores of Control Group 1 (4.1 ± 0.6) and Control Group 2 (4.5 ± 0.7). Evidence from semi-structured interviews indicates that 83.3% of participants in the experimental group acknowledge the superior comprehensibility of structured feedback, which is primarily attributed to the integration of defect type, theoretical rationale, and revision direction into a coherent explanatory chain, rather than the fragmented error notifications provided by conventional tools. For example, when a deficiency related to "argumentative logic discontinuity" is detected, the system associates the feedback with the "argumentative logic–coherence" attribute defined in the ontology and the "claim–dependency–evidence" relational rule, thereby offering traceable theoretical grounding. In contrast, Control Group 1 typically provides only generic prompts such as "logic is unclear," while feedback generated by Control Group 2 lacks support from a unified academic knowledge framework.

Writing quality improvement results. Improvements in writing quality across users with different English proficiency levels are reported in Table 3. The experimental group demonstrates significantly higher values in three-dimensional score improvement, ability transfer rate, and defect correction rate than both control groups, thereby confirming Hypothesis H3.

Table 3. Comparison of writing quality improvement metrics across English proficiency levels

Group	Proficiency Level	Three-Dimensional Score Improvement (mean ± SD)	Ability Transfer Rate (%, mean ± SD)	Defect Correction Rate (%, mean ± SD)	*p* Value vs. ES (Same Level, CG1/CG2)
Experimental Group (ES)	Beginner	2.0 ± 0.3	80.2 ± 3.3	76.3 ± 3.0	–
	Intermediate	2.5 ± 0.2	85.7 ± 2.8	81.2 ± 2.5	–
	Advanced	1.8 ± 0.2	82.0 ± 3.0	78.0 ± 2.7	–
	Overall	2.1 ± 0.3	82.6 ± 3.1	78.5 ± 2.8	–
Control Group 1 (CG1)	Beginner	1.1 ± 0.2	63.5 ± 3.5	60.1 ± 3.3	<0.001/<0.001
	Intermediate	1.3 ± 0.2	66.8 ± 3.2	63.5 ± 3.1	<0.001/<0.001
	Advanced	1.2 ± 0.2	65.6 ± 3.4	63.3 ± 3.2	<0.001/<0.001
	Overall	1.2 ± 0.2	65.3 ± 3.4	62.3 ± 3.2	<0.001/<0.001
Control Group 2 (CG2)	Beginner	1.4 ± 0.2	69.8 ± 3.2	67.2 ± 3.0	<0.001/<0.001
	Intermediate	1.6 ± 0.2	73.5 ± 3.0	70.8 ± 2.8	<0.001/<0.001
	Advanced	1.5 ± 0.2	71.2 ± 3.1	68.1 ± 2.9	<0.001/<0.001
	Overall	1.5 ± 0.2	71.5 ± 3.1	68.7 ± 2.9	<0.001/<0.001

With respect to three-dimensional score improvement, the experimental group achieves an overall gain of 2.1 ± 0.3 points, representing increases of 75.0% relative to Control Group 1 and 40.0% relative to Control Group 2. Among proficiency levels, the most pronounced improvement is observed for intermediate-level users, while gains of 2.0 ± 0.3 points and 1.8 ± 0.2 points are recorded for beginner-level and advanced-level users, respectively. This pattern indicates that progressive feedback effectively aligns with learners' zones of proximal development: foundational language conventions are reinforced for beginners, logical construction is emphasized for intermediate users, and refinement of academic expression is targeted for advanced users, thereby enabling personalized guidance. In terms of ability transfer rate, the experimental group attains 82.6 ± 3.1%, exceeding Control Group 1 by 26.5% and Control Group 2 by 15.5%. These results demonstrate that the system not only corrects immediate writing deficiencies but also facilitates long-term transfer of writing competence. This effect is attributed to the instructional orientation of progressive feedback, in which learners' self-correction and optimization abilities are gradually cultivated through three-tier prompts, rather than through direct rewriting as adopted by conventional tools. Regarding defect correction rate, the experimental group records 78.5 ± 2.8%, which is substantially higher than the values observed for Control Group 1 (62.3 ± 3.2%) and Control Group 2 (68.7 ± 2.9%). This finding further substantiates the pedagogical value of explainable feedback, as a clearer understanding of defect causes and revision logic enables more precise modification behaviors.

4.3 Ablation study: Verification of innovation module contributions

To quantify the contributions of the three core innovation modules and the lightweight mobile adaptation strategy, four ablation variants (AB1–AB4) were constructed. Using the full system as the reference, relative changes in key performance indicators were compared. The results are summarized in Figure 3.

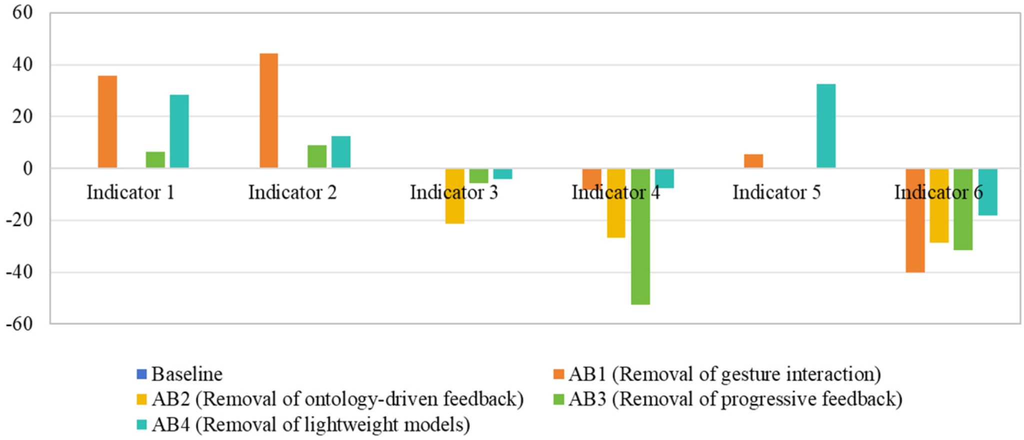

Fig. 3. Comparison of percentage changes in key performance indicators across ablation variants

Indicators 1–6 correspond to the percentage change in task completion time, IOC, feedback precision, writing quality improvement, power consumption, and user satisfaction, respectively. In AB1, the mobile-native gesture interaction paradigm is removed and replaced with conventional click-based and input-field interactions. As a result, task completion time increases by 35.7%, IOC rises by 44.4%, and user satisfaction decreases by 40.2%, with $p < 0.001$. These findings indicate that mobile-native gesture interaction constitutes a core module for ensuring both interaction efficiency and user experience. In AB2, the ontology-driven feedback mechanism is removed and substituted with traditional keyword-matching feedback. Consequently, feedback precision decreases by 21.3%, the explainability score drops to 3.8, and the defect correction rate declines by 19.1%, thereby confirming the decisive role of the ontology knowledge base in sustaining high-quality feedback. In AB3, progressive feedback is removed and replaced with a one-time, comprehensive feedback strategy. Under this configuration, the writing quality improvement value decreases by 52.6%, and the ability transfer rate declines by 28.3%, demonstrating that progressive feedback is critical for achieving instructional empowerment and directly influences gains in writing proficiency. In AB4, local lightweight models are removed, and all assessments are performed exclusively in the cloud. This modification leads to a 32.4% increase in power consumption and a 45.7% increase in interaction latency, indicating that the lightweight adaptation strategy provides essential support for efficient system operation in mobile environments.

5 CONCLUSION

In response to the core challenges of mobile academic English writing, an interactive English writing guidance system integrating natural language processing (NLP) was designed and implemented, and three core innovations were proposed and empirically validated. A mobile-native interaction paradigm integrating text, gestures, and feedback was constructed, overcoming the interaction redundancy inherent in conventional desktop-derived tools. An ontology-based explainable feedback mechanism was designed to address the black-box limitations of existing feedback systems. In addition, a dynamic assessment–integrated progressive feedback generation algorithm was developed to deliver personalized guidance aligned with varying user proficiency levels. Multidimensional comparative experiments confirmed that the proposed mobile-native gesture interaction reduced task completion time to 12.3 minutes, achieved a gesture recognition accuracy of 92.3%, and constrained power consumption to 287 mW, significantly outperforming traditional interaction modes. The ontology-based explainable feedback mechanism attained a feedback precision of 89.7% and an F1 score of 0.86, substantially exceeding the performance of mainstream tools. An SUS score of 82.6 was obtained for the experimental group, and 86.1% of users expressed willingness for long-term use, demonstrating strong contextual adaptability and user acceptance. Results from the ablation study further indicated that removal of mobile-native gesture interaction led to a 35.7% increase in task completion time, removal of the ontology-driven feedback mechanism resulted in a 21.3% reduction in feedback precision, and removal of progressive feedback caused a 52.6% decrease in writing quality improvement. These findings collectively demonstrate that the three core innovation modules function synergistically and together constitute the system's core competitive advantages in mobile academic English writing guidance.

6 REFERENCES

[1] M. Jamshed, A. S. M. M. Ahmed, M. Sarfaraj, and W. U. Warda, "The impact of ChatGPT on English language learners' writing skills: An assessment of AI feedback on mobile," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 19, pp. 18–36, 2024. https://doi.org/10.3991/ijim.v18i19.50361

[2] W. Deng and L. Wang, "The impact of mobile technology on English writing teaching: The relationship between interactive feedback and autonomous learning abilities," *International Journal of Interactive Mobile Technologies*, vol. 19, no. 2, pp. 180–194, 2025. https://doi.org/10.3991/ijim.v19i02.53747

[3] T. H. Brown and L. S. Mbati, "Mobile learning: Moving past the myths and embracing the opportunities," *International Review of Research in Open and Distributed Learning*, vol. 16, no. 2, pp. 115–135, 2015. https://doi.org/10.19173/irrodl.v16i2.2071

[4] E. Baran, "A review of research on mobile learning in teacher education," *Journal of Educational Technology & Society*, vol. 17, no. 4, pp. 17–32, 2014.

[5] C. Aaron, T. Heap, A. Archibald, L. Keyes, M. Novosad, and A. Fein, "Applications of human-AI interaction to optimize teaching workload and improve student writing," *Journal of Computing in Higher Education*, vol. 37, no. 4, pp. 1427–1459, 2025. https://doi.org/10.1007/s12528-025-09474-0

[6] G. Pishtari *et al.*, "Learning design and learning analytics in mobile and ubiquitous learning: A systematic review," *British Journal of Educational Technology*, vol. 51, no. 4, pp. 1078–1100, 2020. https://doi.org/10.1111/bjet.12944

[7] M. Stevenson and A. Phakiti, "The effects of computer-generated feedback on the quality of writing," *Assessing Writing*, vol. 19, pp. 51–65, 2014. https://doi.org/10.1016/j.asw.2013.11.007

[8] J. Wilson, R. Roscoe, and Y. Ahmed, "Automated formative writing assessment using a levels of language framework," *Assessing Writing*, vol. 34, pp. 16–36, 2017. https://doi.org/10.1016/j.asw.2017.08.002

[9] P. H. Hung, G. J. Hwang, Y. F. Lin, T. H. Wu, and I. H. Su, "Seamless connection between learning and assessment-applying progressive learning tasks in mobile ecology inquiry," *Journal of Educational Technology & Society*, vol. 16, no. 1, pp. 194–205, 2013.

[10] R. T. Huang, "Explore the moderating impact of learners' anticipated emotions on mobile learning outcome: A moderated mediation model," *Innovations in Education and Teaching International*, vol. 60, no. 6, pp. 872–882, 2023. https://doi.org/10.1080/14703297.2022.2076717

[11] S. Wang, J. Iwata, and D. Jarrell, "Learning via mobile phones—students' learning styles, needs, preferences and concerns," *International Journal of Innovation and Learning*, vol. 19, no. 4, pp. 431–443, 2016. https://doi.org/10.1504/IJIL.2016.076797

[12] Y. R. Guo and D. H. L. Goh, "Evaluation of affective embodied agents in an information literacy game," *Computers & Education*, vol. 103, pp. 59–75, 2016. https://doi.org/10.1016/j.compedu.2016.09.013

[13] I. Radu, "Augmented reality in education: A meta-review and cross-media analysis," *Personal and Ubiquitous Computing*, vol. 18, no. 6, pp. 1533–1543, 2014. https://doi.org/10.1007/s00779-013-0747-y

[14] M. Sandstead and A. Kibler, "Voice in L2 writing in the age of AI," *Journal of Second Language Writing*, vol. 69, p. 101212, 2025. https://doi.org/10.1016/j.jslw.2025.101212

[15] Y. Zhang, H. Feng, Y. Zhao, and S. Zhang, "Exploring the application of the artificial-intelligence-integrated platform 3D slicer in medical imaging education," *Diagnostics*, vol. 14, no. 2, p. 146, 2024. https://doi.org/10.3390/diagnostics14020146

7 AUTHOR

Jingyi Cai graduated from Henan Normal University with a Master's degree and is currently employed at the School of International Education, Xuchang Vocational and Technical College, focusing on research in the field of English (E-mail: caijingyi1984@163.com).