

ESLA in Senior High School Spoken English Teaching: Dilemmas and Solutions

Jia Liu

School of Foreign Languages, China West Normal University, Nanchong, China

Email: 1915148619@qq.com

How to cite this paper: Liu, J. (2026). ESLA in senior high school spoken English teaching: Dilemmas and solutions. *International Journal of Social Science, Education and Humanities*, 2(2), 61-71. ISSN Print: 3104-4239, ISSN Online: 3104-4247.

<https://doi.org/10.63313/IJSSEH.9037>

Published: 2026-05-25

Copyright © 2026 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

This study focuses on the adaptation issues of ELSA (English Language Speech Assistant), an artificial intelligence speech-assisted tool, in senior high school spoken English teaching. Through a systematic analysis of three dimensions—educational adaptability, technical accuracy, and teaching interactivity—it reveals the core deficiencies of ELSA in the specialized training for the Gaokao spoken English test, classroom management support, and the depth of cultural teaching. The findings indicate that ELSA currently suffers from critical problems such as insufficient adaptation to Gaokao question types, recognition bias of dialectal accents, delayed real-time feedback, and lack of teacher-student interaction, making it difficult to meet the demands of core literacy-oriented senior high school English teaching. To address these problems, this study proposes a multi-level optimization strategy: in terms of educational adaptability, develop a specialized module for the Gaokao spoken English test and enhance the teacher-end management functions; in terms of technical accuracy, construct a "semantic-phonetic" joint model and optimize the real-time feedback system; in terms of teaching interactivity, reconstruct the educational value of gamification and innovate a multimodal integration model. Verified by the strategies, this study provides a theoretical framework for the localized upgrading of ELSA and a practical paradigm for the in-depth integration of AI educational technology in senior high school English teaching, which is of significant reference value for promoting the organic integration of intelligent technology and the cultivation of core competencies in the English discipline.

Keywords

ELSA; Intelligent Speech-Assisted Tool; Spoken English Teaching; Speech Recognition

1. Introduction

With the deep integration of artificial intelligence technology and language education, the application value of intelligent speech-assisted tools in spoken English teaching has become increasingly prominent. Notably, artificial intelligence

technology enables students to conduct self-analysis and summary of knowledge content and learning methods through the introduction of intelligent learning devices and online platforms, thereby optimizing the effectiveness of English teaching (Yang & Yu, 2024). Xu demonstrated how large language models can function in teaching listening, speaking, reading, writing and translation through prompt engineering cases, pointing out that large language models serve as intelligent assistants for English teaching (Xu & Zhao, 2024). As a deep learning-based AI speech recognition application, ELSA (English Language Speech Assistant) has been widely used worldwide thanks to its precise pronunciation correction function. The construction of the English subject competency performance index system has laid a theoretical foundation for the development of English subject competency assessment tools, and provided an operable guidance framework for interpreting assessment results and implementing classroom teaching practices, exerting an important and positive impact on English subject competency assessment, the application of assessment results, classroom teaching reform, and students' learning mode reform (Wang & Chen, 2019). This policy orientation has put forward higher requirements for intelligent speech assessment tools, yet existing technical solutions still have obvious deficiencies in examination standard adaptability and teaching management functions. Existing research shows that current studies on ELSA have two major limitations: first, most research subjects focus on adult learners and higher education, lacking systematic research targeted at the senior high school English teaching scenario; second, most research perspectives pay attention to the technical implementation level, ignoring the in-depth integration of technology and educational scenarios.

This study has both theoretical and practical significance. Theoretically, it constructs a scenario adaptation framework for intelligent educational technology and expands the application of second language acquisition theory in intelligent environments. Practically, it develops an AI-based Gaokao spoken English test scheme, optimizes the educational application of intelligent speech technology, solves the adaptation problems of ELSA, provides a reference for the development of intelligent education, and promotes the development of smart teaching in the era of Educational Informatization 2.0.

2. Defect Analysis

2.1. Defects in Educational Adaptability

2.1.1. Lack of Specialized Training for Gaokao Spoken English Test

ELSA is a pronunciation-focused language learning system built on natural speech rules, international phonetic recognition, and daily dialogue scenarios. While its fundamental training framework enables learners to improve basic pronunciation accuracy, it suffers from a structural deficiency in specialized modules for the exam-oriented Gaokao spoken English test. As a standardized evaluation, the

Gaokao spoken English test assesses more than pure pronunciation proficiency, adopting diverse question types centered on situational Q&A, passage retelling, and topic presentation that require learners' comprehensive language application competence. Taking the Shanghai Gaokao English listening and speaking test as a typical case, its question types cover sentence reading, passage reading, situational questioning, picture description, quick response, and passage-based question answering. This test integrates the assessment of basic pronunciation knowledge and advanced abilities including passage retelling and topic presentation (Xu, 2021). For example, although ELSA's phonetic training enhances learners' single-word pronunciation accuracy, it provides no systematic guidance on discourse-level reading strategies such as sentence stress and sense group segmentation in passage retelling tasks. Consequently, ELSA fails to fully adapt to the current requirements of the Gaokao spoken English test.

2.1.2. Weak Classroom Management Functions

Driven by artificial intelligence and other information technologies, English classroom management has shifted from conventional discipline supervision to systematic management covering learning data monitoring, personalized instruction and group collaboration. For exam-oriented Gaokao oral English teaching, teachers rely on accurate data analysis to conduct tiered teaching and targeted tutoring. Nevertheless, ELSA's teacher-side functions remain limited to basic learning record keeping, falling far short of practical classroom management demands.

Its drawbacks lie in inadequate data refinement and disconnected in-class and after-class training. The system only offers general statistics including class learning time and average pronunciation accuracy, without detailed analysis on individual error patterns and learning weaknesses. Besides, it fails to link in-class teaching with after-class drills. Senior high school English teaching follows a closed loop of lecture, practice and consolidation, yet ELSA cannot assign targeted exercises matching classroom teaching content, making it hard for teachers to consolidate key knowledge via after-class training.

2.1.3. Insufficient Depth of Cultural Teaching

Cross-cultural competence is defined as a key curriculum objective in literacy-oriented senior high school English education, demanding students to perceive Sino-foreign cultural disparities, build cultural cognition and develop multicultural mindsets (Ministry of Education, 2020). ELSA's conversational exercises merely focus on linguistic practice, featuring abundant stereotyped scenarios and superficial language training with in-depth cultural implication and logic neglected. The system fails to fully support cross-cultural teaching, reflected in simplistic cultural settings and scattered cultural explanation. Its dialogues are

confined to routine occasions like dining consumption with rigid interaction patterns, while senior high school teaching requires cross-cultural learning covering academic exchanges and traditional festivals. Furthermore, cultural knowledge on the platform is presented in fragmented form. It only adds brief cultural tips in certain contexts, for instance simply introducing turkey as a typical Thanksgiving food, without systematic elaboration on the festival's origin, cultural connotation and interpersonal etiquette. Such disjointed content hinders students from forming coherent cultural thinking, leaving them unable to grasp essential cultural logic in real cross-cultural communication.

2.2. Defects in Technical Accuracy

2.2.1. Examination-Oriented Deviation of Speech Recognition

Pronunciation, grammatical accuracy and content coherence are core scoring criteria of Gaokao oral English test, requiring learning platforms to integrate phonetic, grammatical and logical training. Nevertheless, ELSA's automatic speech recognition engine is predominantly designed for pronunciation assessment, making its training targets inconsistent with exam standards. The limitation lies in weak grammatical error identification and vacant discourse logic evaluation. The system fails to effectively detect non-phonetic errors, frequently overlooking or misjudging subject-verb disagreement, tense errors and improper conjunction use in oral tests. Besides, it lacks relevant analytical functions. Since passage retelling and topic presentation demand capabilities in information arrangement, logical linking and argument construction, ELSA cannot conduct valid assessment on these discourse competence.

2.2.2. Insufficient Adaptability to Dialectal Accents

In the Chinese senior high school English teaching environment, dialectal accents exert a significant impact on students' English phonetics acquisition, yet the ELSA system shows obvious insufficient adaptability in coping with this problem. This deficiency severely restricts its teaching application effect nationwide, especially in regions with prominent dialect characteristics. Viewing from the influence of Chinese dialects on English pronunciation, dialects in different regions have unique characteristics at the phonetic level, which will transfer to English pronunciation and form typical dialectal accent problems. For example, the confusion of "n/l" is common in Sichuan dialect, leading to high school students in Sichuan prone to pronunciation errors when pronouncing English words such as "night" and "light"(Xiao,2014). However, the speech model of ELSA has obvious limitations in design and cannot effectively deal with the pronunciation problems caused by Chinese dialects. The model is mainly trained based on non-standard accents such as those in India and Southeast Asia. Although the accents in these regions also have their uniqueness, they are quite different from the influence of Chinese dialects on

English pronunciation. This deviation in training data results in the insufficient ability of ELSA to recognize and correct Chinese dialectal accents.

2.2.3. Lag of Real-Time Feedback

In the group training scenario of senior high school English classrooms, the accuracy and timeliness of real-time feedback are key factors to ensure the effect of spoken English training. Nevertheless, the ELSA system has a significant lag problem in this regard. ELSA's real-time feedback has a delay of about 0.8 seconds, which disrupts students' follow-up reading rhythm and exerts a substantial impact on the mastery of key phonetic skills such as liaison and weak reading. The negative effect of this problem is more prominent in question types with high requirements for rhythm and phonetic coherence such as "passage reading". From the perspective of actual classroom teaching operation, senior high school English teachers often adopt the training mode of "group follow-up reading—real-time correction", requiring students to imitate immediately after hearing the standard pronunciation to form correct phonetic muscle memory. However, the feedback delay of the ELSA system breaks this ideal training rhythm. After students finish follow-up reading and pronunciation, the system needs to wait for about 0.8 seconds to give feedback. This time difference prevents students from comparing their pronunciation with the standard sound in a timely manner, making it difficult to adjust their pronunciation methods in the first place.

2.3. Defects in Teaching Interactivity

2.3.1. Defects in Motivation Maintenance Mechanism

Sustained learning interaction depends on students' intrinsic participation motivation, and relevant research indicates that students' learning engagement gradually declines from junior to senior high school due to increasing curriculum difficulty and frequent learning setbacks, with insufficient motivational support in learning activities serving as the primary cause (Guo et al., 2025). Nevertheless, the ELSA system possesses prominent flaws in motivating and sustaining students' learning engagement. First, its misplaced teaching orientation undermines interactive motivation: overemphasizing mechanical pronunciation correction of tone and intonation while neglecting the essential communicative nature of oral English, the system reduces interactive learning to passive task completion rather than enabling students to recognize the value of oral interaction for opinion and emotion expression. Second, its simplistic feedback mechanism dampens learning enthusiasm: ELSA's evaluation system merely adopts pronunciation accuracy and practice duration as core assessment metrics and ignores the creative value of language output, failing to satisfy senior high school students' psychological needs for recognition and in-depth communication and ultimately inducing long-term interactive fatigue and learning resistance. For example, when students practice speeches on themes such as "My Future Career", ELSA is only capable of assessing

the pronunciation of key words rather than evaluating the logical rigor and emotional appeal of students' viewpoints.

2.3.2. Lack of Teacher-Student Interaction

Interaction originates from the inherent communicative attribute of language and is an objective requirement and necessary condition for second language acquisition (Loewen & Sato, 2018). However, the excessive technical application of ELSA leads to the loss of communicative interaction between teachers and students in traditional classrooms. First, when ELSA becomes the main interactive carrier, the teacher's role degenerates from an "interaction guide" to a "technical operator". The face-to-face real-time communication in traditional classrooms (such as impromptu questioning in classroom discussions and roaming guidance in group activities) is replaced by fragmented program instructions, weakening the emotional resonance and thinking resonance carried by eye contact and body language between teachers and students. Second, the in-depth absence of psychological interaction scenarios: senior high school students originally have "error anxiety" in spoken English expression. If ELSA's "precise error correction" lacks synchronous emotional support from teachers (such as fault tolerance and thinking guidance), it will further aggravate students' psychological pressure in spoken English training, leading students to tend to "speak less to make fewer mistakes" rather than take the initiative to output in a safe interactive atmosphere. At the same time, it essentially cuts off the collaborative training chain of "language knowledge—emotional attitude—thinking quality". Students fail to truly express emotional attitudes in mechanical practice, let alone achieve the cultivation of thinking quality. Although students participate in language activities, they do not really construct the ability and willingness to communicate interactively in English.

3. Optimization Strategies

3.1. Improving Educational Adaptability

3.1.1. Development of Specialized Module for Gaokao Spoken English Test

To address ELSA's structural deficiencies in targeted Gaokao oral English training, developing exam-specialized modules is critical to enhancing the platform's educational adaptability. It is necessary to integrate official Gaokao oral question resources to build targeted training modules for situational Q&A, passage retelling, and viewpoint presentation. The situational Q&A module can set up diverse themes covering campus life and social hotspots with typical examples to cultivate students' situational expression and logical thinking abilities. The passage retelling module, while retaining basic pronunciation training, should prioritize systematic guidance on discourse skills including stress placement, sense group division and logical connection by adopting authentic Gaokao passages that are segmented and marked with standardized pauses and stresses to help students master retelling skills

through imitation and practice. The viewpoint presentation module can adopt frequently tested Gaokao topics such as environmental protection and technological development to train students in viewpoint construction, argument organization and cohesive expression. Furthermore, the integration of localized Gaokao scoring criteria into ELSA enables precise, exam-aligned performance evaluation, allowing students to conduct targeted training that conforms to regional assessment requirements and improves their exam preparation efficiency.

3.1.2. Strengthening of Teacher-End Functions

To tackle existing data management issues, an integrated classroom management system needs to be established. It enables bulk import of students' personal and academic information to support dynamic ability-based grouping and differentiated task distribution, such as targeted learning packages for underachieving learners. Enhanced error analysis can identify pronunciation mistake categories and generate personal weakness profiles, while class-wide error statistics assist teachers in adjusting teaching focus and forming a closed loop covering data monitoring, instructional adjustment and targeted practice. Meanwhile, boundaries between offline teaching and online learning should be eliminated. Functions including real-time screen sharing, group competition assessment and voice message interaction help teachers review learning outcomes instantly, organize interactive oral challenges and respond to students' on-site questions. These diversified interactive features convert ELSA from a mere after-class practice tool into an integral classroom teaching component, solidifying the learning cycle of classroom instruction, in-class drills and after-school consolidation.

3.1.3. Expansion of the Depth of Cultural Teaching

Against ELSA's shallow cultural training limited to linguistic exercises, a three-dimensional cultural teaching module shall be built based on senior high school English core literacy requirements for cross-cultural competence. Beyond daily conversational settings, high-context scenes involving academic exchanges and festival activities will be supplemented. Taking Thanksgiving dialogue as an example, apart from basic food-related expressions, animated clips and illustrations can interpret the cultural evolution of traditional food and family dining etiquette, helping learners link linguistic forms with cultural implications. Meanwhile, textbook-aligned interactive courses can convert passages about paper-cutting and traditional Chinese medicine into role-play tasks of introducing native culture. Equipped with a database covering seasonal origins, agricultural relevance and modern cultural application, the system facilitates students' in-depth cultural interpretation during oral practice, which fits classroom teaching schedule and effectively promotes learners' cultural cognition via interactive learning.

3.2. Optimization of Technical Accuracy

3.2.1. Optimization of Examination-Oriented Speech Recognition

To optimize the examination-oriented deviation of ELSA in senior high school spoken English teaching, it is of great significance to introduce a "semantic-phonetic" joint model. Referring to the General Senior High School English Curriculum Standards (2017 Edition, 2020 Revision), the Gaokao spoken English test has clear requirements for students' grammatical accuracy and content coherence (Gao, 2025). In this context, combining the BERT model with the ASR engine is a feasible approach. Based on the pre-trained language model, BERT can analyze the logical coherence of students' spoken responses through context and accurately identify grammatical problems such as subject-verb disagreement and tense errors; the ASR engine focuses on phoneme correction, and the two cooperate to achieve multi-dimensional evaluation of pronunciation, grammar and content logic, in line with the Gaokao scoring standards.

3.2.2. Upgrading of Real-Time Feedback System

In view of the 0.8-second feedback delay in classroom group training, edge computing technology can be adopted to deploy speech processing nodes on the school's local server. This scheme shortens the data transmission path and avoids network delay in cloud processing, reducing the feedback time to within 0.3 seconds. Taking "passage reading" training as an example, the system can respond in real time when students follow up reading, ensuring that the pronunciation rhythm of liaison, weak reading and so on is synchronized with the original sound, effectively solving the problem of rhythm chaos caused by delay.

Develop a visualized dynamic pronunciation feedback module to present the accuracy of students' pronunciation in the form of a heat map: mark the deviation areas of easily confused phonemes such as "th" and "n/l" in red, and standard pronunciation intervals in green. Teachers can adjust the key points of guidance in real time according to the heat map. For example, if students are found to be weak in pronouncing the "th" sound, targeted tongue twister practice can be designed; students can also intuitively see their pronunciation defects through the chart. For example, students in Sichuan dialect areas often confuse "n/l", and can compare with the standard pronunciation position with the help of the heat map, forming a learning closed loop of "real-time feedback—precise correction".

3.3. Enhancement of Teaching Interactivity

3.3.1. Reconstruction of Educational Value of Gamification Design

Oral training can be designed as role-play games, such as simulating UN climate conferences that require learners to apply complex grammar and academic vocabulary, with authentic conference videos and policy documents accessible after tasks to realize integrated learning of language use and knowledge expansion. Team cooperation modes can also be adopted. As an effective innovative teaching

approach to enhance students' overall ability and classroom interaction, group learning gains wide application in modern education (Yan, 2024). Students can carry out themed debates and cross-cultural planning projects, finishing data sorting, argument building and oral presentation collaboratively, while the system tracks their linguistic performance and logical thinking to boost critical thinking and speaking confidence. Furthermore, a dual incentive mechanism combining academic achievements and social recognition can be set up. Task completion earns medals and redeemable points for tutoring resources, and class rankings and community sharing functions can further sustain students' learning enthusiasm.

3.3.2. Innovation of Multimodal Teaching Integration

Advancing technology drives educational innovation, and multimodal teaching has been widely applied across disciplines (Shao, 2024), which facilitates students' English knowledge acquisition in senior high school teaching. A connected training system can integrate listening, speaking and reading modules. Taking the TED speech *How to Speak so that People Want to Listen* as material, students listen to the audio, summarize core communication principles and interact with AI tutors, and read relevant passages to accumulate thematic vocabulary, achieving comprehensive ability advancement. Additionally, the film dubbing and script adaptation module enables scenario-based creative practice. Students dub classic movie clips and receive assessment on pronunciation and emotion, then rewrite scripts into campus stories with complex grammar structures. Such learning from imitation to independent creation improves phonetic proficiency and creative language competence.

3.3.3. Simulation of Teacher-Student Interaction Scenarios

In view of students' fear of difficulties in learning, optimize the dynamic guidance mechanism of AI foreign teachers and implement teacher-like guidance strategies for AI foreign teachers. When students make frequent mistakes or remain silent for a long time in practice, the system will automatically push encouraging sentences such as "We often practice this pronunciation difficulty in class, take your time", and synchronously adjust the task difficulty—such as splitting complex long sentences into short sentences for follow-up reading, completing single sentence grammar practice first and then integrating into complete dialogue, simulating the layered teaching thinking of teachers in class.

The whole process also needs to give play to the supervisory role of teachers. Therefore, a three-party linkage interactive interface of teachers, students and AI is developed, allowing teachers to view the dialogue between students and AI in real time through the management background. When finding that students deviate from arguments or make expression errors in the "environmental protection theme debate", teachers can directly intervene in the dialogue for comments, such as "The

data citation just now is very accurate, and adding XX cases will be more convincing", and insert voice explanation clips into the AI dialogue process, forming a dual-teacher teaching mode of "AI basic training + teacher personalized guidance".

4. Conclusion

This study systematically analyzes the adaptation dilemma of ELSA in senior high school spoken English teaching and reveals the structural contradictions in the application of artificial intelligence speech tools in educational scenarios. The research shows that to realize the transformation from a "pronunciation correction tool" to an "intelligent teaching system", ELSA must break through three core barriers: complete the shift from general training to examination specialization in functional design, realize the leap from single speech recognition to multi-dimensional language evaluation in technical architecture, and construct a new relationship from human-computer dialogue to teacher-student collaboration in interaction mode. The optimization strategy system proposed in this study provides a feasible scheme for the localized transformation of ELSA through specific paths such as the development of Gaokao modules, optimization of dialect models, and multimodal integration. These schemes not only solve the current technical adaptation problems, but also promote the organic integration of AI tools and teaching practice through the design of strengthening teacher-end functions and expanding the depth of cultural teaching. Future research can further explore the integration path of large language models and speech technology, develop intelligent systems with teaching decision support functions, and pay attention to data ethics and educational equity in technology application. The theoretical framework and practical strategies of this study have enlightening significance for the innovation of the cultivation mode of core competencies in the English discipline in the artificial intelligence era, and provide an important reference for the development of smart teaching under the background of Educational Informatization 2.0.

References

- [1] Yang, W., & Yu, P. (2024). Research on the current situation and countermeasures of artificial intelligence empowering college English classroom teaching evaluation. *Journal of Hubei Open Vocational College*, 37(8), 139–141.
- [2] Xu, J. J., & Zhao, C. (2024). The role of large language models in English teaching. *Foreign Language Education Research Frontiers*, 7(1), 3–10, 90.
- [3] Wang, Q., & Chen, Z. H. (2019). Measurement, evaluation and teaching improvement of English subject competence under the background of core literacy. *China Examinations*, (3), 13.
- [4] Xu, W. (2021). Practice of oral assessment in large-scale high-stakes examinations—A case study of Shanghai college entrance examination English listening and speaking test. *Foreign Language Testing and Teaching*, (1), 21–27.
- [5] Ministry of Education of the People's Republic of China. (2020). General high school English curriculum standard (2017 ed., revised 2020). People's Education Press.

-
- [6] Xiao, C. C. (2014). A brief discussion on the influence of Sichuan dialect phonetic characteristics on English pronunciation. *English Teachers*, 14(11), 68-72.
- [7] Guo, J. C., Zheng, Y., Wang, X. Z., et al. (2025). Promotion strategies of daily academic motivation resilience among middle school students. *Teaching and Administration*, (12), 66-71.
- [8] Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, 51(3), 285-329.
- [9] Gao, R. J. (2025). Problems sorting and practical exploration of senior high school oral English teaching. *Ningxia Education*, (Z1), 136-137.
- [10] Yan, Y. F. (2024). Effective practice of group cooperative learning in junior high school English classrooms. *English Campus*, (24), 115-117.
- [11] Shao, C. L. (2024). Application of multimodal teaching mode in junior high school English teaching. *Education Circle*, (31), 26-28.