ISSN 2522-6398 Vol. 9, Issue 2: 125-131, DOI: 10.25236/FER.2026.090216

Research on Domestic Tablets in AI-Integrated English Listening and Speaking Teaching and Examinations

Wang Jinsong

Wuhan Optics Valley No.1 Junior Middle School, Wuhan, 430071, China
15893428@qq.com

Abstract: *Against the backdrop of the rapid development of educational informatization and the steady advancement of the domestic substitution strategy, this paper introduces a fully domestic tablet computer constructed based on the domestic RISC-V architecture K1 chip and the Open Harmony 5.0 operating system. It details the innovative application methods of AI-integrated human-machine dialogue for English listening and speaking teaching and examinations carried out in four schools: Wuhan Optics Valley No.1 Junior Middle School, The First Affiliated Middle School of Central China Normal University (Junior High Division), The Affiliated Middle School of Tongji Medical College of Huazhong University of Science and Technology, and Wuhan Meijia Foreign Language School. Furthermore, it summarizes the advantages of these tablets compared to traditional school computer rooms in terms of adaptability to classroom teaching scenarios, improvement of examination efficiency, and assurance of test paper equivalence. Through this practical research, the study aims to provide a replicable implementation paradigm for the innovative application of information technology in the education sector and contribute an empirical case with promotional value to the national domestic substitution strategy for new educational infrastructure.*

Keywords: *Domestic chips; HarmonyOS; English listening and speaking; Domestic substitution; Junior high school academic proficiency test*

1. Introduction

Under the guidance of the Compulsory Education English Curriculum Standards (2022 Edition) and the 2024 new textbooks, English education is undergoing a pivotal shift toward practical communication, necessitating innovations in smart, AI-driven pedagogy. This transformation aligns with nationwide assessment reforms, where over 20 provinces have integrated listening and speaking into high-stakes exams via AI-supported "human-machine dialogue."[1] However, in the context of global technological competition, reliance on foreign underlying technologies for these systems poses significant security risks. Consequently, adhering to national mandates for "domestic substitution" has become urgent. Developing secure, localized educational infrastructure is now essential to support these dual reforms in teaching and assessment while safeguarding national data security.

2. Problems Existing in Traditional Computer Room Models

Research indicates that computer labs are the primary venue for English listening and speaking assessments in China. While AI-integrated scoring systems improve efficiency, the limited number of labs creates a "many students, few labs" bottleneck, leading to significant implementation challenges.

2.1. Problems in the Teaching Process

From a teaching perspective, the primary function of computer labs is to meet the requirements of the school's Information Technology curriculum. If English listening and speaking instruction requires frequent occupation of the labs, schools must constantly adjust the class schedule to fit the labs' idle slots. [2] This makes it difficult to conduct English listening and speaking lab courses with any regularity or consistency. Secondly, it is difficult to integrate the lab environment with the ordinary classroom environment. This necessitates scheduling listening and speaking classes separately, disrupting the continuity of normal teaching. Students must switch learning scenarios at different times and locations,

ISSN 2522-6398 Vol. 9, Issue 2: 125-131, DOI: 10.25236/FER.2026.090216

consuming time and energy. [3]Moreover, the closed environment of the computer lab lacks the natural atmosphere of daily English learning, limiting the innovation of teaching methods and the improvement of teaching effectiveness. Furthermore, the noise-blocking partitions installed in computer labs make it difficult for teachers to comprehensively monitor every student's computer operations during class. For instance, it is common for students to use computers for non-class-related activities, such as being distracted, browsing the web, or playing games. Such behaviors disrupt classroom order and directly weaken the quality of English listening and speaking instruction.

2.2. Problems in the Examination Process

Generally, schools have an average of 1-2 computer labs, and the upper limit of usable device points is fixed. Therefore, exam administration requires arranging candidates into multiple batches to queue for the exam. For example, if 500 students participate in a test and the lab holds a maximum of 50 people at a time, the exam must be divided into 10 batches, resulting in long duration and low efficiency. During multi-batch exams, to prevent earlier batches from leaking test content to later batches, it is necessary to prepare test papers that are similar in difficulty but different in content for different batches. [4] However, in practice, achieving "test paper equivalence" is highly difficult. Since it is hard to ensure complete consistency across different papers in terms of vocabulary coverage, topic difficulty, and task complexity, guaranteeing exam fairness is challenging. In addition, the need for multi-batch exams and multiple sets of test papers increases the workload and time required for test construction, grading, and proctoring. During the test construction phase, teachers need more time to write and review multiple sets of different papers; during grading, differences in answers and scoring standards for each set increase complexity and time costs.

3. Application Methods of Tablet Computers

Compared with traditional computer lab desktops, tablet computers offer the advantages of high integration and easy maintenance. In terms of flexibility, as portable devices, tablets are not restricted by the space of dedicated scenarios. They can be flexibly deployed and used across various classes, achieving integration with classroom teaching and exam scenarios in ordinary classrooms. This paper introduces the specific application methods of using domestic tablet computers to conduct AI-integrated junior high school English listening and speaking teaching and exams, as practiced by the author in four schools: Optics Valley No.1 Junior Middle School, No.1 Middle School Affiliated to Central China Normal University (Junior High Division), The Middle School Attached to Tongji Medical College of HUST, and Wuhan Mejia Foreign Language School.

3.1. Classroom

Teaching Application Leveraging the portability and intelligent interactive features of tablet devices, this study fully utilized the schools' additional English listening and speaking classes, as well as fragmented time slots such as morning reading, lunch breaks, and evening self-study. Tablets were distributed in the classroom, and under the supervision and guidance of teachers, daily teaching applications were conducted, including English vocabulary recitation and dictation, text reading and recitation, and classroom exercises. These activities were integrated into the daily curriculum, achieving an organic combination of information technology and daily teaching.

3.1.1. Vocabulary Recitation and Dictation

Based on the requirements of the Compulsory Education English Curriculum Standards (2022 Edition) for the four-dimensional capabilities of unit core vocabulary—"Pronunciation, Form, Meaning, and Usage"—there is still a significant implementation gap in current teaching. Surveys show that traditional vocabulary dictation models focus mostly on spelling accuracy and matching Chinese definitions, while the assessment of pronunciation standards and pragmatic ability is weak. This leads to a fragmented phenomenon in students' vocabulary learning: "high dictation accuracy but low pragmatic conversion rate." To address this, a "Three-Stage Closed-Loop Vocabulary Dictation Method" was designed based on information technology.

Schools distribute tablets and dedicated headphones in classrooms during fragmented periods like morning reading or breaks to conduct 15-20 minute vocabulary dictation sessions.

Stage 1: "Sound-Form Mapping." The system plays standard pronunciation, and students must

ISSN 2522-6398 Vol. 9, Issue 2: 125-131, DOI: 10.25236/FER.2026.090216

complete the dictation within a time limit. Upon completion, they take a photo and upload it. The AI engine corrects it via OCR recognition and spelling database comparison, synchronously generating a heatmap of error-prone words.

Stage 2: "Meaning-Usage Association." The system randomly shuffles the word order and requires students to complete three interactions: (1) Select the correct definition from a dynamic word cloud (containing synonym distractors); (2) Vocabulary reading test (with optional phonetic hints or hidden mode), where the speech recognition module assesses pronunciation accuracy in real-time (down to phoneme-level deviation detection).

Stage 3: "Adaptive Reinforcement." Based on data from the first two stages, the system automatically pushes customized consolidation exercises—embedding new words into situational dialogues and matching them with fill-in-the-blank drills.

Compared to traditional paper-and-pencil dictation, the tablet intelligent dictation model presents four structural advantages:

Multi-modal Input Verification: Integrates speech recognition, image recognition, and natural language processing technologies to achieve a full-chain capability assessment of "Audio Input - Text Output - Speech Production - Contextual Application."

Dynamic Knowledge Graph: The system automatically constructs an individual vocabulary mastery matrix, visually presenting the sound-form association strength and meaning-usage conversion efficiency for each word.

Formative Assessment System: Each training session generates a diagnostic report containing pronunciation waveform comparisons, spelling error pattern analysis, and pragmatic quality radar charts, enabling teachers to precisely pinpoint common class weaknesses.

Neuroscience Reinforcement Mechanism: Based on the spaced repetition algorithm designed around the Ebbinghaus Forgetting Curve, the long-term memory retention rate of vocabulary increased from 28% in the traditional model to 67% ($p<0.01$). Experimental data shows that in unit vocabulary assessments for classes using this model, the excellence rate in pronunciation (≥90 points) jumped from 19% to 74%, the context application score rate increased by 41%, and the slope of the forgetting curve decreased by 62% after two weeks.

3.1.2. Text Reading and Recitation

Currently, when students preview or recite texts, due to a lack of effective technical means, methods such as teacher spot checks, group recitation, peer checking, or parents recording videos to upload to WeChat groups are commonly used. Surveys indicate the following primary problems: (1) It is common for students to recite mechanically with a "task-completion" mindset. The "laundry list" style of recitation is prominent, lacking attention to pronunciation details such as aesthetic appeal, intonation, and linking/reduction, leading to the fossilization of pronunciation errors. (2) The rate of forgetting recited texts is high; previously recited content is often forgotten after subsequent teaching begins. (3) Inspection methods struggle to achieve comprehensive coverage. Teachers and parents, limited by time and energy, can hardly conduct precise, word-for-word reviews, leading to slip-ups and students making up numbers.

Therefore, in this experiment, an optimized text recitation system was designed using tablet teaching software and IT means, adopting the following innovative methods: (1) AI Speech Assessment Intervention: The system scores text content sentence by sentence, with precision down to the word level. It marks the student's pronunciation with different colors: Content Error (Red), Non-standard Pronunciation (Yellow), and Correct Recitation (Green), providing immediate feedback. Students can click on problematic parts to hear correct pronunciation demonstrations and definitions, correcting their oral issues. (2) Role-Play Mode: For situational dialogues and multi-character passages, a role-play mode is introduced. It allows for Human-Machine cooperation or Peer-to-Peer cooperation, with role-switching options to enhance situational immersion and increase students' interest in reciting. (3) Recitation Leaderboards: A top-3 leaderboard is established based on completion time and scoring to incentivize students to complete recitations within a specified time, using gamification to stimulate efficiency and interest. (4) Smart Prompting: Unlike the past where students would flip through the book when stuck, the system identifies the recitation progress. When a student gets stuck, it can prompt the next 1-2 words as a guide and mark it as a "forgetting point." In subsequent recitations, these points are displayed as hints and then hidden once mastery is achieved, improving efficiency.

ISSN 2522-6398 Vol. 9, Issue 2: 125-131, DOI: 10.25236/FER.2026.090216

By deeply integrating cognitive neuroscience with intelligent technology, a closed learning ecosystem of "Diagnosis - Training - Reinforcement - Assessment" is constructed, providing a quantifiable digital solution for language learning. Experimental data shows that students using the intelligent recitation system improved in dimensions such as fluency scores, recitation speed, and pronunciation accuracy compared to the traditional teaching group.

3.1.3. Classroom Listening and Speaking Application

Based on the core competency guidance of the Compulsory Education English Curriculum Standards (2022 Edition) and the design philosophy of the 2024 PEP new junior high school English textbooks, this study explores innovations regarding the practical difficulties of cultivating language output ability in unit teaching. [5]The new textbooks explicitly set oral expression requirements such as "talk," "introduce," "discuss," and "present" in the "Project" section of each unit. However, traditional classrooms, constrained by time, space, and a lack of personalized guidance, often struggle to achieve precise oral instruction. To this end, this study constructs an "AI + Smart Classroom" teaching model. It extracts content related to oral expression within the unit and, relying on tablets and intelligent teaching platforms, converts the language ability requirements in the curriculum standards into observable, quantifiable digital literacy indicators through a closed-loop system of "Data Collection - Intelligent Diagnosis - Virtual Training - Multi-dimensional Evaluation." Notably, this model innovatively integrates cross-cultural communicative competence and technology application competence, forming a "Language Ability +" comprehensive training framework.

Taking the teaching practice of Grade 7, Unit 3 "My School" as an example, this unit aligns with the "Man and Society" theme of the curriculum standards, focusing on spatial description and campus culture cognition. The teaching design adopts a four-stage progressive structure:

(1) Intelligent Diagnosis Stage (10 minutes) Using a speech recognition system built on deep learning algorithms, an AI reading diagnosis is first conducted. The teacher pushes a customized corpus containing target sentence structures (This is... / There is... / We can... here, etc.) to student terminals. The system uses ASR (Automatic Speech Recognition) technology for multi-dimensional analysis. Beyond standard pronunciation accuracy, parameters such as logical stress (e.g., emphasizing prepositions of place) and fluency (e.g., spatial description sequence) are specifically added. The generated personalized diagnostic report includes: (1) Pronunciation Heatmap; (2) Intonation Waveform Comparison; (3) Excellent Example Library. The teacher uses an interactive electronic whiteboard for visual commentary, focusing on deconstructing the discourse structure of spatial descriptions.

(2) Virtual Training Stage (15 minutes) Based on situated cognition theory, an MR (Mixed Reality) campus scene is constructed. The system pushes a virtual campus built with 3D modeling tools to the tablets, generating a semantic map containing geographic coordinates and building functions. In AI role-play dialogues, the system uses NLP (Natural Language Processing) technology to act as a "New Student NPC" for voice interaction and dynamic feedback: (1) Spatial Logic Detection: Triggers a prompt when the student's description order conflicts with the map's topological structure; (2) Function Match Analysis: Uses word vector models to judge semantic associations like "library - quiet place"; (3) Cross-Cultural Contrast: When a student mentions "playground," the system pushes real-scene videos of similar places in British/American schools for cognitive expansion.

(3) Extended Application Stage (10 minutes) Adopting the CLIL (Content and Language Integrated Learning) concept, the documentary Schools Around the World is selected for multi-modal input. The intelligent dubbing system features: (1) Voice-Subtitle Sync Analysis: Uses Dynamic Time Warping (DTW) algorithms to assess dubbing alignment; (2) Cultural Annotation Generation: Automatically identifies cultural symbols in the video (e.g., British uniforms, American lockers) and generates bullet-screen annotations; (3) Collaborative Learning Support: When synthesizing group recordings, the system uses voiceprint recognition technology to assess individual contribution levels.

(4) Metacognitive Improvement Stage (5 minutes) The teacher-side system integrates data from the entire process to generate a class capability matrix, focusing on: (1) Accuracy of preposition use (e.g., data shows 76% of students confuse 'in/on'); (2) Frequency of cultural references (only 12% of students spontaneously contrasted Chinese and foreign campuses); (3) Based on this, targeted guidance is provided, and tiered extension tasks are assigned.

ISSN 2522-6398 Vol. 9, Issue 2: 125-131, DOI: 10.25236/FER.2026.090216

3.2. Application in Formal Exams

The administrative process for tablet-based exams aligns with traditional models but excels in flexibility and adaptability. By utilizing standard classrooms and contact-based data transmission, it removes the reliance on fixed computer labs. The system allows for dynamic scheduling, supporting simultaneous testing across multiple rooms or multi-batch rotations to significantly increase testing capacity.For multi-batch implementations (recommended max 3 batches), a three-zone system (Waiting, Preparation, and Exam Rooms) with strict unidirectional flow is required. To ensure fairness and securely reuse the same test papers, all batches should be completed within 90 minutes. Rigorous management of batch transitions is critical to preventing candidate contact and test leakage.

3.2.1. Exam Room Layout

Regarding equipment requirements, each exam station must be equipped with one tablet computer and one dedicated exam headset. Each exam room also requires one proctor tablet and a tablet storage/charging unit. Seats in the exam room should be arranged individually, with a spacing of no less than 600mm to ensure sufficient space between candidates. Exam headsets should feature unidirectional sound pickup, good noise cancellation, and status indicator lights. They must also possess data storage capabilities to prevent interference between answering voices and ensure candidates can hear audio recordings clearly. Additionally, schools with standardized exam room conditions should activate signal shielding and video surveillance systems during the exam to ensure security and fairness.

3.2.2. Data Distribution and Collection

Each exam room is configured with one proctor tablet for use by invigilators. Before the exam, test papers and administrative information are pre-loaded onto the proctor tablet. 30 minutes prior to the exam start, the proctor switches the tablet to "Test Paper Distribution" status and uses contact-based transmission (touching devices) to securely transfer exam data to each candidate's tablet one by one. After the exam begins, the proctor switches to "Data Collection" status. Once each batch concludes, the proctor again touches each exam tablet to complete data retrieval. The proctoring system automatically ensures all answer data is complete before collecting it into the proctor tablet. Afterward, all proctor tablets are returned to the exam administration office and connected to the management computer to aggregate and export the answer packages from all rooms.

3.2.3. Exam Implementation

During the exam, candidates must complete a sound test to ensure the device is functioning normally. Once the exam starts, the system software automatically jumps to the answering interface. Candidates select answers for objective questions using touch and record answers for oral questions using the headset microphone, following system prompts. Upon completion, the exam software generates an answer package.

3.2.4. Grading and Scoring

After answer packages are collected and imported into the intelligent grading system, the system implements differentiated scoring strategies based on question types. For objective questions in the listening section (e.g., multiple-choice, fill-in-the-blank), fully automated and precise scoring is performed based on preset standard answers. For oral questions, a three-stage "Expert Calibration - Machine Learning - Intelligent Scoring" model is adopted. First, a team of language assessment experts manually scores typical speech samples to construct a standardized scoring framework. Next, deep learning algorithms extract features from expert scoring standards to train models, forming an intelligent scoring engine capable of semantic understanding, pronunciation assessment, and fluency analysis. Finally, the system uses the trained AI model to automatically score all oral responses. To ensure scientific and reliable results, a multi-engine parallel scoring mechanism can also be used during grading. This involves cross-verification by integrating scoring engines with different algorithmic logic. When the score difference between engines exceeds a preset threshold, the system automatically triggers an arbitration procedure, where professional human graders review and adjudicate the disputed responses. This hybrid model of "Intelligent Primary Scoring + Human Auxiliary Verification" fully leverages the high efficiency of AI while effectively controlling scoring deviations through human-machine collaboration.

ISSN 2522-6398 Vol. 9, Issue 2: 125-131, DOI: 10.25236/FER.2026.090216

4. Advantages of Domestic Chips and Operating Systems

This study uses the scenario of English listening and speaking human-machine dialogue teaching and exams as an entry point to explore the innovative application value of tablet computers based on the domestic RISC-V architecture "K1" chip and the open-source HarmonyOS 5.0 operating system. Experiments have proven that these domestic tablet devices build the following core advantages through deep software-hardware collaborative optimization: First, scenario-based system customization empowers precise teaching. Leveraging the high extensibility of open-source HarmonyOS 5.0, the system reconstructs the audio processing framework specifically for voice interaction scenarios, achieving ultra-low latency in voice collection and real-time feedback. The K1 chip's customized RISC-V instruction set supports parallel acceleration of voice feature extraction and AI noise reduction algorithms. This boosts the response speed of oral assessment beyond traditional x86 and ARM architectures, ensuring smooth operation of high-frequency dialogue training in class. Second, a "Pure Mode" developed for teaching and exams. Through underlying operating system controls, unauthorized access is blocked to ensure process fairness and prevent candidate misoperations from affecting the normal conduct of the exam. Third, a full-stack domestic foundation ensures safety and controllability. The K1 chip uses a completely autonomously designed 22nm RISC-V microarchitecture, paired with HarmonyOS 5.0's self-developed distributed microkernel. This forms a full-link security barrier from the chip instruction set to the OS kernel, effectively defending against man-in-the-middle attacks and risks of data leakage or tampering. Fourth, independent technological innovation drives superior performance. The K1 chip innovatively adopts a heterogeneous computing architecture with 2 TOPS of independent AI computing power. Combined with HarmonyOS 5.0's deterministic latency engine and lightweight thread scheduling algorithms, the device remains stable and efficient during long-term operation, significantly outperforming similar Android devices. Fifth, ergonomic optimization enhances the teaching experience. Capitalizing on the energy efficiency of the 22nm K1 chip and HarmonyOS 5.0's adaptive refresh rate technology, battery life is extended to 10 hours, meeting all-day teaching needs. A specially developed visual voice feedback system displays pronunciation issues in real-time via 3D voiceprint spectrograms, greatly improving human-machine interaction efficiency. Empirical evidence from pilot schools shows that this solution offers significant efficiency gains when applied to English listening and speaking classrooms and exams.

5. Conclusion

This study provides a regional reference for education informatization by constructing an application model based on domestic chips and operating systems. However, current results reflect a specific technological ecosystem with three limitations: sample bias toward high-informatization schools, failing to represent rural contexts ; a lack of longitudinal data beyond Grade 7 and restriction to the 2024 PEP textbooks ; and reliance on traditional data collection that restricts granular analysis. Future research should focus on multi-dimensional mechanisms: establishing comparative frameworks across diverse development zones, implementing K-12 longitudinal tracking to understand technological adaptation, and utilizing advanced behavioral analysis systems for data aggregation. Ultimately, achieving breakthroughs through interdisciplinary intersection and multi-method integration is essential to fully unleashing the revolutionary potential of intelligent educational equipment.

References

[1] Wang Jinsong, He Xin.(2024). Design and Implementation Concept of Junior Middle School English Listening and Speaking Teaching Model Integrating Artificial Intelligence. Experiment and Equipment for Primary and Secondary Schools, , (01).

[2] Ministry of Education, Central Cyberspace Affairs Commission, National Development and Reform Commission, Ministry of Industry and Information Technology, Ministry of Finance, & People's Bank of China. (2021). Guidance on promoting the construction of new educational infrastructure and building a high-quality education support system. Retrieved February 11, 2025, from https://www.beijing.gov.cn/zhengce/zhengcefagui/qtwj/202204/t20220407_2656309.html

[3] Wang, P. (2020). Analysis of the role of intelligent listening and speaking training systems in improving junior high school students' English listening and speaking abilities. Exam Weekly, (89).

[4] Chou, M.-H. (2016). Strategy use for listening in English as a foreign language: A comparison of academic and vocational high school students. TESOL Journal, (3).

ISSN 2522-6398 Vol. 9, Issue 2: 125-131, DOI: 10.25236/FER.2026.090216

[5] Chou, M. H. (2019). The impact of the English listening test in the high-stakes national entrance examination on junior high school students and teachers. International Journal of Listening.