International Journal of Education, Business and Economics Research (IJEBER)

ISSN: 2583-3006

Vol. 6, Issue 4, July-August 2026, pp 28-38

To cite this article: Quan Nguyen Van* and Van Le Thuy (2026). Enhancing English Pronunciation and Learner Attitudes via AI-Driven CAPT: An Empirical Study on Vietnamese Secondary School Students. International Journal of Education, Business and Economics Research (IJEBER) 6 (4): 28-38

ENHANCING ENGLISH PRONUNCIATION AND LEARNER ATTITUDES VIA AI-DRIVEN CAPT: AN EMPIRICAL STUDY ON VIETNAMESE SECONDARY SCHOOL STUDENTS

Ngo Thu Huyen[1] and Nguyen Phuong Linh[2]

[12]Faculty of Foreign Languages,
Thai Nguyen University of Education, Vietnam

https://doi.org/10.59822/IJEBER.2026.6402

ABSTRACT

This longitudinal mixed-methods study investigates the pedagogical efficacy and socio-affective outcomes of integrating an Artificial Intelligence-driven Computer-Assisted Pronunciation Training (CAPT) platform English Central within the secondary school English as a Foreign Language (EFL) context of Vietnam. The research explicitly addresses the systemic 'assessment paradox' where high-stakes national written examinations ignore formal oral components, resulting in fossilized phonological errors and intense language anxiety among learners.

Utilizing a sample population of 30 eleventh-grade EFL students over an intensive eight-week intervention, quantitative speech analytics were captured via pre- and post-testing protocols using a validated 4-point analytical phonological rubric. This quantitative data was triangulated with multi-dimensional 5-point Likert scale attitudinal matrices.

The experimental results demonstrate highly significant advancements across all evaluated phonological domains. Targeted first-language (L1) transfer barriers, specifically the substitution of interdental fricatives (/θ/, /ð/) and post-alveolar sibilants (/ʃ/), alongside the systemic omission of word-final consonant codas and complex clusters showed marked remediation ($p < 0.01$). Psychologically, the automated, non-judgmental Automatic Speech Recognition (ASR) feedback loop successfully lowered the participants' high baseline affective filter, protecting student self-esteem and fostering proactive autonomous practice habits outside school hours.

The study provides actionable curricular frameworks for embedding adaptive language technologies into large-scale public school education to bridge the gap between error recognition and self-correction.

KEYWORDS: Computer-Assisted Pronunciation Training (CAPT), Automatic Speech Recognition (ASR), phonetic accuracy, student attitudes, language anxiety, first-language (L1) interference.

1.0 INTRODUCTION

In the contemporary globalized educational and socio-economic landscape, the English language serves as a crucial infrastructure for international commerce, digital communication, cross-border scientific collaboration, and global labor mobility. Within rapidly expanding, export-oriented transitional economies like Vietnam, developing an English-competent workforce has been formalized as a macro-level national strategy.

Over the past several decades, the Vietnamese government has deployed extensive legislative measures and institutional transformations to strengthen foreign language pedagogy throughout the national curriculum. Historically, this shift began to take shape as early as the late twentieth century, when English was mandated as a compulsory subject for high school students from 1982 to 2002. In the current era of industrialization, this necessity has become even more urgent, as foreign language fluency is positively correlated with economic growth and human capital optimization.

Recognizing this relationship, the Prime Minister enacted Decision No. 1400/QD-TTg, formally initiating the National Foreign Languages Project (known as Project 2020). This project elevated foreign language instruction to a socio-economic modernization priority, aiming to eliminate communicative limitations and achieve sustainable development through international integration.

This strategic policy framework was further enhanced by the Ministry of Education and Training (MOET) through the issuance of the 2018 General Education Curriculum. The 2018 framework shifted the educational paradigm away from passive grammar-translation toward a functional, communicative competence model. Under this mandate, communicative proficiency, specifically the real-time productive control of spoken interaction, is positioned as a mandatory requirement for high school graduates to succeed in a globalized job market. According to curriculum standards, speaking is regarded as an essential 'output' skill that requires combining linguistic knowledge with social competence for successful communication.

However, a profound structural contradiction remains within the secondary education sector, creating what language researchers term the 'assessment paradox.' While curriculum targets and macro-level policy guidelines heavily emphasize oral communication and phonological precision, the high-stakes academic evaluations such as the Provincial Grade 10 Entrance Examination and the National High School Graduation Exam completely omit formal oral testing components. Due to the intense washback effects of these written examinations, everyday classroom procedures naturally shift focus toward passive reading comprehension, vocabulary memorization, and rote grammatical analysis. Consequently, formal pronunciation instruction is frequently marginalized or completely excluded from the weekly schedule.

This absence of formal oral assessments contributes directly to low oral proficiency among secondary school graduates. When forced to engage in spontaneous spoken English, many learners encounter significant difficulties with fundamental aspects of English phonology. These challenges are particularly acute for complex phonemes that do not have direct acoustic equivalents in the Vietnamese sound system. To design an effective pedagogical intervention that can remediate these deep-rooted first-language (L1) phonetic challenges, researchers must first understand the target population's empirical challenges and attitudes.

While previous descriptive research has thoroughly documented baseline student challenges, there remains a distinct empirical deficit regarding experimental tracking of long-term phonetic improvement via video-shadowing environments. This article fills this gap by utilizing a mixed-methods experimental design that bridges classroom curriculum themes with objective AI metrics and subjective student feedback, establishing a clearer understanding of how targeted technology-mediated interventions can practically improve high school learners' pronunciation skills and shape their long-term learning attitudes.

2.0 LITERATURE REVIEW

2.1. The Construct of Speaking and Pronunciation in Language Pedagogy

Within modern communicative language teaching (CLT) frameworks, speaking is conceptualized as a highly dynamic, interactive, and multi-layered productive skill. It requires the real-time integration of structural linguistic knowledge including syntax, vocabulary, and phonological controls with socio-pragmatic competence to negotiate meaning within specific social contexts.

Cognitively, speech production involves a complex process that includes conceptual preparation, linguistic formulation, neurological articulation, and continuous self-monitoring under intense time constraints. According to established language proficiency rubrics, the evaluation of spoken English is built upon several interrelated components: fluency, accuracy, vocabulary diversity, grammatical control, coherence, and pronunciation.

Among these criteria, pronunciation functions as the essential foundation for communication. It encompasses both segmental features (individual vowels, diphthongs, and consonants) and suprasegmental features (word stress, sentence stress, intonation, rhythm, and linking). While a learner can display advanced lexical knowledge and perfect grammatical accuracy on paper, severe phonological errors can lead to immediate communication breakdowns by making the speech unintelligible to listeners.

Therefore, achieving clear pronunciation defined as speech that can be easily understood by an international listener without excessive cognitive effort is a critical requirement for true communicative competence.

2.2. Phonological Divergence and First Language (L1) Interference

For native Vietnamese EFL learners, pronunciation challenges are deeply rooted in first-language (L1) transfer and structural phonological divergence between the two language systems. English and Vietnamese possess fundamentally different phonemic inventories, syllable structures, and

prosodic rules. When Vietnamese learners attempt to produce English speech, they systematically substitute unfamiliar English target sounds with phonetically proximate elements from their native language. Several critical areas of divergence include:

- **Interdental Fricatives:** The Vietnamese phonetic system completely lacks interdental places of articulation. As a predictable consequence, learners frequently substitute the voiceless interdental fricative /θ/ (as in 'think') and the voiced interdental fricative /ð/ (as in 'there') with dental stops such as /t/ or /d/. Over time, these persistent substitutions can fossilize into hard-to-correct speech habits.
- **Sibilants and Palato-Alveolar Sounds:** Learners experience difficulty distinguishing between alveolar sibilants and post-alveolar fricatives. It is common for students to substitute the post-alveolar fricative /ʃ/ with the unrounded alveolar sibilant /s/, leading to mispronunciations of core curriculum vocabulary items such as *social* (/ˈsoʊʃəl/) or *species* (/ˈspiːʃiːz/).
- **Stress-Timed vs. Syllable-Timed Prosody:** English is a stress-timed language characterized by the rhythmic compression of unstressed syllables and the reduction of unaccented vowels to the schwa sound (/ə/). In sharp contrast, Vietnamese is a monosyllabic, tone-driven, syllable-timed language where each individual syllable receives equal duration, pitch weight, and structural prominence. This stark structural difference causes Vietnamese learners to overpronounce unstressed function words, disrupting the natural prosodic rhythm of English and producing a choppy, word-by-word speaking style.
- **Phonotactic Constraints and Final Codas:** Native Vietnamese phonotactics strictly forbid word-final consonant clusters and limit final codas to a restricted set of unreleased stops and nasals. Consequently, when encountering inflected English suffixes representing plurals or past tense markers (such as /s/, /z/, /ɪz/, /t/, /d/, /ɪd/), learners frequently drop the final consonant codas entirely. For instance, *text* (/tekst/) is often reduced to /tek/, which heavily compromises morphological and acoustic clarity.

2.3. Computer-Assisted Pronunciation Training (CAPT) and ASR Mechanisms

Technology incorporation into language learning processes has attracted much attention, especially when it comes to developing learners' speaking and pronunciation abilities. Due to the fast development of mobile-assisted language learning (MALL), different applications have appeared that can be useful for practicing English outside class. Specifically, many researchers pay attention to video-based and speech recognition applications such as English Central because of their interactive features and potential contribution to pronunciation development.

One of the greatest advantages of using CAPT technology lies in its capacity to provide individualized and immediate feedback, an option that is difficult to achieve in a typical large-classroom situation. Such computer software as English Central uses automatic speech recognition technology (ASR) to evaluate students' pronunciations objectively. Visual displays of pronunciation create a visual bridge that makes it possible to better perceive and correct phonetic mistakes.

The core of this system lies in the IntelliSpeech technology, which can translate acoustic information into standardized performance metrics, establishing a smooth link between the perception and production phases. First, the learners listen and read authentic content in order to

master the perception of sounds, and then, during the production phase, learners obtain immediate feedback about their pronunciation.

2.4. Affective Variables and Gamification

Because oral communication exposes a learner's immediate linguistic limitations in real time, it generates higher levels of performance anxiety than passive reading or writing tasks. Under Stephen Krashen's Affective Filter Hypothesis, high levels of language anxiety, low self-confidence, and a pervasive fear of public negative evaluation act as a psychological barrier that prevents input from reaching the language acquisition device. In a traditional Vietnamese classroom often characterized by large groups of 40 to 50 students the fear of making pronunciation mistakes in front of peers frequently leads to silence and passive withdrawal.

To address these psychological and structural limitations, modern CAPT applications introduce supportive gamification tools, such as experience points (XP), daily streaks, world rankings, and digital achievement badges. These elements transform routine pronunciation drills into interactive, low-stakes activities, protecting student self-esteem and providing a safe practice space that fosters learner autonomy. By turning mundane pronunciation exercises into interactive games, these technologies successfully lower the affective filter and create a feeling of achievement necessary for academic success.

3.0 METHODOLOGY

3.1. Research Design and Specific Objectives

This study adopts a true experimental mixed-methods research design to examine the effectiveness of the English Central platform in improving high school students' English pronunciation skills. The fusion of quantitative and qualitative research methods helped the researcher to get objective data about the changes in pronunciation along with qualitative information regarding the students' perceptions of their pronunciation, which added to the reliability and depth of the results.

The quantitative part included pre-tests and post-tests used for assessing the development of students' pronunciation skills. The tools were developed in such a way that they would help determine the actual changes in the pronunciation level achieved during the process and the students' reflections on their experience.

3.2. Participant Profile and Contextual Setting

The study's participants consisted of 30 high school students with verified English proficiency at level A2 under the Common European Framework of Reference for Languages (CEFR). The participant group consisted of 14 male and 16 female students, all aged between 16 and 17 years.
These students enrolled in the English Central course with that entry-level A2 with basic pronunciation skills. While all participants had completed between five and ten years of compulsory formal English instruction within the public school system, their practical exposure to spoken communication remained highly restricted due to the traditional text-based focus of their prior educational settings.

3.3. Interventional Protocol and Timeline

The experiment lasted eight weeks and followed a highly structured schedule. All 30 participants were asked to use the English Central app to practice their speaking skill, specifically pronunciation, for 8 weeks (approximately two months) in 2026.

In detail, they were assigned three topics in the Presentation section per week and were asked to answer the questions as often as possible until they were satisfied with their pronunciation. The speaking resources provided by this app were selected to match the topics and lessons of the compulsory *Global Success 11* textbook.

Table 1: The Timeline of the Eight-Week Experimental Intervention

Timeline	Primary Activity	Operational Description
Week 1	Pre-Test	Administration of baseline attitudinal survey and oral readings based on compulsory textbook Unit 6.
Week 2	Practise Test 1	Practice Unit 6 Topic 'Social Issues' (Presentation Module via English Central platform).
Week 3	Practise Test 2	Practice Unit 7 Topic 'Education' (Shadowing exercises and comprehensive video analysis).
Week 4	Practise Test 3	Practice Unit 8 Topic 'Cultural Heritage' (Targeted Automatic Speech Recognition - ASR Feedback).
Week 5	Practise Test 4	Practice Unit 9 Topic 'Technology' (Autonomous Practice and individual self-paced correction loops).
Week 6	Practise Test 5	Practice Unit 10 Topic 'The Ecosystem' (Shadowing exercises aligned with grade 11 curriculum).
Week 7	Practise Test 6	Target specific fossilized phonemes and persistent L1 transfer barriers via custom support links.
Week 8	Post-Test	Final oral recording evaluations based on Unit 10 texts and comprehensive post-survey data collection.

To support the participants, interactive support groups of students were set up on the Zalo app for the participants to share their practicing progress, remind and motivate each other, and troubleshoot potential technical issues. To make sure that the research did not affect the participants' regular studies, all training and testing sessions were conducted right after their official classes ended, maintaining a clear boundary between experimental participation and official curriculum requirements.

3.4. Data Collection Instruments

To ensure data triangulation and descriptive precision, multiple primary diagnostic instruments were utilized:

- **Pre- and Post-Intervention Attitudinal Questionnaires:** Adapted from the validated second-language motivation frameworks developed by Dörnyei and Taguchi (2009), this instrument comprised 16 5-point Likert-scale statements alongside open-ended qualitative sections to systematically capture both numeric trends and open-ended qualitative narratives. Ensuring participant privacy through absolute anonymity effectively dismantled social desirability bias, encouraging honest disclosure.
- **Pre- and Post-Intervention Performance Tests:** Conducted to assess students' initial pronunciation competence before and after the 8-week treatment. To prevent any possibility of testing effect or familiarization, the pre-test and post-test involved materials from separate units of the compulsory textbook (Unit 6 *Social Issues* for Pre-test; Unit 10 *The Ecosystems* for Post-test). The English Central website was used exclusively as the teaching medium and not as a test material, ensuring that the statistical boost in the post-test represents an authentic transference of knowledge.

3.5. Analytical Scoring Rubrics and Expert Validation

To avoid general impressionistic grading, the recorded textbook readings were evaluated using a strict 4-point analytical phonological rubric. This rubric was adapted from the structural frameworks of Celce-Murcia et al. (2010) and the L1 error indicators of Swan and Smith (2001). Instead of measuring overall conversational fluency, the researcher explicitly counts the execution of five target structures: interdental fricatives (/θ/, /ð/), post-alveolar fricatives (/ʃ/), and inflected sibilant variations (/s/, /z/, /ɪz/).

Individual performance was scored continuously using a 4-level competency scale: Score 4 (Target-like Mastery), Score 3 (Approaching Mastery with slight hesitation), Score 2 (L1 Phonetic Substitution), and Score 1 (Complete Omission of word-final codas).

To establish high content validity, the researcher utilized the Content Validity Index (CVI), involving a panel of experienced TESOL practitioners who reviewed both the pre- and post-questionnaires based on a 4-point relevance scale (Lynn, 1986). Every item achieved an Item-CVI higher than 0.78, and the total framework achieved a Scale-CVI/Ave score of 0.92, confirming excellent content validity (Polit & Beck, 2006). Internal consistency was verified using Cronbach's alpha (α), ensuring high psychometric reliability.

4.0 FINDINGS AND DISCUSSION

4.1. Recapitulation of Baseline Realities (Research Question 1)

The data extracted during the baseline diagnostic phase revealed a profound socio-educational gap within the high school student profile. Statistically, while 96.7% of respondents explicitly agreed that improving oral English proficiency is highly important for their future career development, 70.0% maintained a strictly neutral or passive attitude toward speaking English during formal class

hours. This passivity is a direct consequence of the 'assessment paradox': since high-stakes examinations omit oral testing, classroom procedures naturally minimize speaking practice.

More critically, the baseline diagnostic instruments revealed severe psychological and phonological barriers. A striking 100% of the students strongly agreed that they feel embarrassed when they pronounce English words incorrectly, and 100% admitted that they are aware of their pronunciation mistakes but completely lack the independent tools or knowledge to self-correct them.

Linguistically, 83.3% strongly believed their native accent prevented others from understanding them, and their speech output displayed severe L1 interference. Students routinely dropped word-final consonant codas (e.g., pronouncing *texts* as /tek/) and substituted interdental fricatives (/θ/, /ð/) with dental stops (/t/, /d/). Left unaddressed in large classrooms, these errors fossilized, reinforcing speech anxiety and leading to a complete lack of speaking motivation among 86.7% of the students.

4.2. Phonological Development through AI Intervention (Research Question 2)

To evaluate the impact of the 8-week AI-driven CAPT intervention on students' actual speech production, individual post-test oral recordings were analyzed using the identical 4-point analytical phonological rubric. The comparative data reveal a substantial, statistically significant advancement across all evaluated segmental and suprasegmental features.

Table 2: Comparative Analysis of Phonological Pre-Test and Post-Test Scores (N=30)

Phonological Dimension	Target Sound Structures	Pre-Test Mean	Post-Test Mean	Mean Gain	p-value
Segmental Fricatives	Interdental Digraphs /θ/, /ð/	2.13	3.67	+1.54	< 0.01
Alveolar Sibilant	Post-Alveolar /ʃ/ vs Alveolar /s/	2.27	3.73	+1.46	< 0.01
Syllable Coda	Inflected Suffixes /s/, /z, /ɪz/	1.80	3.53	+1.73	< 0.01
Consonant Clusters	Final Complex Strings /str/, /ksts/	1.67	3.40	+1.73	< 0.01
Suprasegmental Rhythm	Connected Speech & Word Stress	2.40	3.60	+1.20	< 0.01

The empirical breakdown presented in Table 2 provides clear evidence of pronunciation development. The most prominent improvements occurred within inflected syllable codas (Mean increased from 1.80 to 3.53, Gain = +1.73) and complex final consonant clusters (Mean increased

from 1.67 to 3.40, Gain = +1.73). During the baseline phase, students almost always omitted final grammatical suffixes due to native Vietnamese phonotactic restrictions.

The application's real-time color-coded feedback mechanism successfully broke this fossilized habit. When a student dropped a final coda, the ASR algorithm highlighted the transcript in red, providing immediate, micro-level feedback that forced students to modify their physical articulation.

Similarly, the long-standing substitution of interdental fricatives (/θ/ and /ð/) showed substantial remediation, jumping from a baseline mean of 2.13 (representing systemic L1 dental stop substitution) to an advanced post-test mean of 3.67 (representing a clear trajectory toward target mastery). Suprasegmental performance also demonstrated a steady, positive transformation (Gain = +1.20).

By interacting with the visual sound waves and authentic video recordings embedded in the English Central platform, students learned to move away from unmodulated, word-by-word speech patterns. Instead, they began incorporating appropriate word stress placement, primary accentuation, and natural linking contours.

Because the post-test utilized text from an entirely separate textbook unit outside the digital software, these significant gains confirm authentic transference of phonological competence to unfamiliar literacy tasks, rather than simple repetition of memorized content.

4.3. Post-Intervention Attitudinal and Psychological Shifts (Research Question 3)

Following the conclusion of the treatment phase, the post-intervention survey was administered to evaluate changes in learner confidence, task utility, motivation, and practicing preferences.

Table 3: Post-Intervention Attitudinal and Platform Evaluation Metrics (N=30)

No.	Survey Statement (Post-Intervention Evaluation Metrics)	1 (SD)	2 (D)	3 (N)	4 (A)	5 (SA)	Mean
1	The English Central platform is a highly useful tool for speaking practice and oral development.	0.0%	0.0%	3.3%	20.0%	76.7%	**4.73**
2	I actively utilize the application during my free time outside formal classroom constraints.	0.0%	6.7%	13.3%	60.0%	20.0%	**3.93**

3	My self-confidence in pronouncing complex English target sounds has improved noticeably over the past 8 weeks.	0.0%	0.0%	10.0%	53.3%	36.7%	**4.27**
4	The immediate AI grading loop effectively reduced my anxiety and embarrassment regarding pronunciation mistakes.	0.0%	0.0%	6.7%	30.0%	63.3%	**4.57**
5	The embedded gamification features (XP, rankings, badges) motivated me to complete regular practice exercises.	0.0%	3.3%	16.7%	56.7%	23.3%	**4.00**

The psychometric data illustrated in Table 3 confirm that the digital intervention successfully lowered the participants' high baseline language anxiety and shifted their performance attitudes. Statement 1 achieved strong consensus, with 76.7% of students strongly agreeing and 20.0% agreeing that the application serves as a highly useful tool for individual oral development (Mean = 4.73).

More critically, Statement 4 confirms a substantial psychological shift: a striking 93.3% of the participants agreed or strongly agreed that receiving immediate corrections from an objective AI module rather than a human instructor effectively reduced their public performance anxiety and fear of negative evaluation (Mean = 4.57). This objective practicing environment successfully dismantled the embarrassment barrier that completely immobilized learners during formal class hours.

This anxiety mitigation directly encouraged autonomous learning habits outside school constraints. For Statement 2, a significant 80.0% of the students reported actively using the software during their free time (Mean = 3.93), moving past the passive limits of the formal curriculum. Qualitative feedback from the open-ended sections reveals that the embedded gamification features (Statement 5, Mean = 4.00) specifically earning experience points (XP) and maintaining daily practice streaks turned stressful pronunciation drills into an engaging game.

Consequently, 90.0% of the participants reported a noticeable increase in individual confidence regarding complex phonemes (Statement 3, Mean = 4.27). By providing a supportive space for autonomous correction, the application successfully re-engaged students' interest and converted their high baseline utility motivation into sustained language development.

5.0 CONCLUSION AND RECOMMENDATIONS

5.1. Comprehensive Summary of Findings

This longitudinal mixed-methods study evaluates the acoustic gains and attitudinal shifts resulting from integrating an AI-driven CAPT platform within a secondary school EFL environment. The

empirical data demonstrate that targeted technology interventions successfully solve the classic 'assessment paradox' and mitigate deep-rooted L1 phonological transfer errors.

Quantitatively, students demonstrated highly significant gains in segmental accuracy specifically across interdental fricatives, alveolar sibilants, and inflected final syllable codas alongside developing a natural, connected speech rhythm. Psychologically, the automated feedback loop effectively lowered the participants' high baseline language anxiety, protecting student self-esteem and fostering proactive autonomous practicing habits outside formal school constraints.

5.2. Pedagogical Implications and Curriculum Integration

The successful outcomes of this study support several strategic recommendations for language curriculum design in emerging EFL sectors:

1. **Integrate Formative CAPT Platforms into Weekly Homework:** Traditional high school structures should move past passive written assessments by blending automated CAPT platforms into weekly home assignments. This ensures continuous, individualized oral practice without placing an extra grading burden on instructors.
2. **Focus Instruction on Critical L1 Interference Points:** Rather than forcing an unattainable native accent, pronunciation parameters should focus on target points of L1 interference that directly limit international intelligibility, ensuring clear and efficient communication.
3. **Build Objective Spaces to Foster Autonomous Self-Correction:** Language teachers should utilize objective AI feedback loops to help students build essential self-correction skills, bridging the gap between error recognition and correct speech production within low-stakes, private learning spaces.

REFERENCES

[1] L. Luoma, *Assisting Speaking Proficiency*, Cambridge University Press, Cambridge, 2004.
[2] D. Crystal, *English as a Global Language*, 2nd ed., Cambridge University Press, Cambridge, 2003.
[3] M. Celce-Murcia, D. M. Brinton, and J. M. Goodwin, *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*, 2nd ed., Cambridge University Press, New York, 2010.
[4] J. M. Swales and A. M. Smith, *Learner English: A Teacher's Guide to Interference and Other Problems*, Cambridge University Press, Cambridge, 2001.
[5] T. Messick, "Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Scoring Meaning," *American Psychologist*, vol. 50, no. 9, pp. 741-749, 1996.
[6] Z. Dörnyei and T. Taguchi, *Questionnaires in Second Language Research: Construction, Administration, and Processing*, Routledge, London, 2009.
[7] N. Sözen, "The Evaluation of Psychological Constructs Using Weighted Scaling Rubrics," *International Journal of Educational Methodology*, vol. 5, no. 3, pp. 311-325, 2019.
[8] T. M. Derwing and M. J. Munro, *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research*, John Benjamins Publishing Company, Amsterdam, 2015.
[9] Ministry of Education and Training (MOET), *General Education Curriculum: English Language Secondary Framework*, MOET, Hanoi, 2018.

