# The Efficiency of an Instructional English Pronunciation Package in a CAPT System for Undergraduate Students: The Integration of Artificial Intelligence and a Human Instructor

# ประสิทธิภาพชุดการสอนการออกเสียงภาษาอังกฤษในระบบคอมพิวเตอร์ช่วยฝึกการออกเสียง สำหรับนักศึกษาในระดับปริญญาตรี: การผสมผสานระหว่างปัญญาประดิษฐ์และครูผู้สอน

Wanvipha Hongnaphadol*[1] Attapol Attanak[2]
วรรณวิภา หงส์นภาดล*[1] อัฐพล อัฐนาค[2]

wanvipha.h@ku.th*

**Received:** March, 16 2022 **Revised:** May, 06 2022 **Accepted:** May, 12 2022

## Abstract

The aims of this study were to: 1) investigate the overall efficiency of the instructional pronunciation package, 2) compare the students' achievement in pronunciation before and after the training, and 3) evaluate the students' satisfaction with the instructional English pronunciation package in a Computer Assisted Pronunciation Training (CAPT) system. The samples were 30 undergraduate students from two public universities, who had been selected by simple random sampling. The pre-experimental research, a one group pretest and posttest design, was conducted with 12 interventions over 4 weeks on the Microsoft Teams platform. The instruments were as follows: 1) the RP application and the IR tool, 2) a pronunciation performance test consisting of a 30-word list for pre-test and post-test, 3) a questionnaire to survey students' satisfaction toward the instructional pronunciation package in the CAPT system, and 4) reading texts of the pronunciation practices, which noted the score profiles at the end of each of the 12 interventions. The data was analyzed with SPSS version 20 to determine Mean, Percentage, Standard Deviation, $E_1/E_2$, and the t-test. The findings revealed that the instructional pronunciation efficiency had been 89.32/86.80, according to the specified 80/80 criteria, reflecting that it had been efficient. The undergraduate students' pronunciation skills had also improved given that the mean score of the post-tests was higher than the pre-test with the level of significance at 0.01. The students' satisfaction with the instructional English pronunciation package in the CAPT system had been strongly positive.

**Keywords:** Instructional Pronunciation Package, Automatic Speech Recognition (ASR), Artificial Intelligence

[1] Lecturer in the Faculty of Management Sciences, Kasetsart University Sriracha Campus
[2] Lecturer in the Language Institute, Khon Kaen University
[1] อาจารย์ คณะวิทยาการจัดการ มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา
[2] อาจารย์ สถาบันภาษา มหาวิทยาลัยขอนแก่น

**บทคัดย่อ**

การศึกษานี้มีวัตถุประสงค์เพื่อ 1) ตรวจสอบประสิทธิภาพของชุดการสอนการออกเสียง 2) เปรียบเทียบผลสัมฤทธิ์ในการออกเสียงของนักศึกษาก่อนและหลังการทดลอง และ 3) ประเมินความพึงพอใจของนักศึกษาต่อชุดการสอนการออกเสียงภาษาอังกฤษในระบบการใช้คอมพิวเตอร์ช่วยการฝึกออกเสียง (CAPT) กลุ่มตัวอย่างเป็นนักศึกษาในระดับปริญญาตรี 30 คน จากมหาวิทยาลัยของรัฐสองแห่ง ซึ่งคัดเลือกโดยการสุ่มตัวอย่างแบบง่าย การวิจัยก่อนการทดลองเป็นแบบแผนการทดลองแบบกลุ่มเดียวทดสอบก่อนหลัง ดำเนินการด้วยการฝึก 12 ครั้งใน 4 สัปดาห์บนแพลตฟอร์ม Microsoft Teams เครื่องมือ ได้แก่ 1) แอปพลิเคชัน Reading Progress และเครื่องมือ Immersive Reader 2) แบบทดสอบประสิทธิภาพการออกเสียงก่อนและหลัง 30 คำ 3) แบบสอบถามเพื่อสำรวจความพึงพอใจของนักศึกษาที่มีต่อชุดการสอนการออกเสียงในระบบ CAPT และ 4) ข้อความอ่านฝึกการออกเสียงโดยระบุคะแนนที่ส่วนท้ายของการฝึกทั้ง 12 ครั้ง วิเคราะห์ข้อมูลโดยใช้ค่าเฉลี่ย ร้อยละ ส่วนเบี่ยงเบนมาตรฐาน $E_1/E_2$ และ t-test โดยใช้โปรแกรมสำเร็จรูปทางสถิติ SPSS 20.0 ผลการวิจัยพบว่า ประสิทธิภาพการสอนการออกเสียงเท่ากับ 89.32/86.80 เป็นไปตามเกณฑ์ 80/80 ที่กำหนด สะท้อนว่ามีประสิทธิภาพ ทักษะการออกเสียงของนักศึกษาพัฒนาขึ้นเนื่องจากคะแนนเฉลี่ยของทักษะการออกเสียงหลังการทดสอบสูงกว่าก่อนการสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และความพึงพอใจของนักศึกษาต่อชุดการสอนการออกเสียงภาษาอังกฤษในระบบ CAPT โดยภาพรวมอยู่ในระดับพึงพอใจมาก

**คำสำคัญ:** ชุดการสอนการออกเสียง, การรู้จำเสียงอัตโนมัติ (ASR), ปัญญาประดิษฐ์

## Introduction

The growth of technological development has spurred the expansion of digital media for foreign language education, which has resulted in new techniques and tools that have proven to be effective in making major contributions to L2 teaching, learning, and research, and in particular, to the analyses of the soundness of pronunciation. Many technological developments have been engaged with Computer Assisted Pronunciation Training (CAPT), such as automatic speech recognition (ASR) and speech analysis. More advanced tools are supported by Artificial Intelligence (AI), which has been developed by holding fast to the principles of how the human brain works. ASR, together with speech analysis technology, are presently empowering applications on mobile phones, such as *Duolingo* or *ELSA Speak*, which are able to give personalized feedback and opportunities for structured communication (Pennington & Rogerson-Revell, 2019).

The success of CAPT is intricately related to several factors: 1) feedback and 2) the learners' notice. Wong and Young (2014) investigated the ASR-based CAPT's different types of feedback: 1) the first level of feedback: pronunciation scores and waveforms; 2) the second level of feedback: comments, lists of correctly and incorrectly pronounced words, and the replay tool for learners, which allowed the students to re-listen to their pronunciation; and 3) the third level of feedback: the accurate word pronunciation demonstrated as isolated words and as words embedded in sentences. The experimental group, which had received all levels of feedback forms, exhibited better pronunciation, while the control group, which had obtained only the first level feedback, complained that the provided waveforms had been difficult to understand. In contrast to the results from a study conducted by Maghrebi, Heydarpour, and Shalmani (2016), it was found that the experimental group, which had received waveforms and pitch contours (considered as the first level feedback) via the *Rosetta Stone* software, had outperformed the control group, which had received a placebo training. This may have resulted because the participants in the study by Maghrebi et al. (2016) may have been visual learners, who had a better ability to process visual information. Olson (2014) also supported the fact that visual feedback, such as spectrograms, can help to spot and to rectify segmental errors. Essentially, teachers should provide a variety of feedback that can meet the learning styles of L2 learners.

All of the given feedback or auditory inputs are not completely converted into intakes by L2 learners (Coder, 1967). In L2 acquisition, inputs become intakes when the learners notice them (Schmidt, 2001). Añorga and Benander (2015) required native English speakers, who were taking Spanish classes to accomplish the following: 1) to record their voices as they read aloud, 2) to compare their voices with the model voices provided, and 3) to reflect upon the comparisons between the voices. This led to pronunciation changes of the target phonemes. Martin (2020) employed Cued Pronunciation Readings (iCPR), which included both perceptual and production training to improve the pronunciation of the L1 English learners of L2 German. The learners in the experimental group were required to discriminate between the accented sounds and the native sounds and in doing so, the significant changes in the target sounds were revealed. Moreover, the experimental group outperformed the control group.

Language learning pedagogies are not perfectly compatible with the affordances of digital technology (Pennington & Rogerson-Revell, 2019). For instance, a CAPT tool might not meet instructional requirements (Rogerson-Revell, 2021). Native-like or near-native pronunciation has

been the ultimate goal, as well as the dominant focus in L2 pronunciation teaching and learning (Jenkins, 2000). Rather than focusing on the principle of nativeness and lessening the accented pronunciation, the aim of teaching pronunciation has shifted and now concentrates on the principles of intelligibility and comprehensibility by equipping L2 learners with the ability to communicate intelligibly with both native and NNSE (Levis, 2005). Nevertheless, the majority of present ASR-based CAPT cannot recognize a variety of utterances from different speakers, especially from the non-native ones (Henrichsen, 2021), and this factor results in poor recognition rates. Compared to native speakers, ASR, which is embedded in many computer programs, decreases the recognition rates to near 70% when processing advanced non-native speakers with foreign accents (Levis, 2007). This may occur because ASR compares and contrasts the auditory inputs with the vocal database of native-speakers (Pennington & Rogerson-Revell, 2019).

The reliability and validity of automated scoring is questionable. The pronunciation scores, given by the human raters, have been found to deviate from the ASR-generated scores. CAPT with insufficient ASR may be annoying and may have unwanted effects on leaners, such as demotivation (Rogerson-Revell, 2021). Moreover, the tendency of giving feedback on pronunciation errors is to provide binary feedback by simply informing learners whether or not their pronunciation is accurate (Henrichsen, 2021; Rogerson- Revell, 2021). Based upon this, students are not necessarily given the assistance they need to fix their pronunciation errors. Several comments have been made concerning the use of AI technology in reading instruction and the accuracy of the pronunciation scores generated by ASR: 1) the computerized voice of the AI tool was not expressive enough when reading since the intonation sounded robotic, and 2) there was a lack of expression and intonation, both of which are required for effective reading instruction (Jarke et al., 2020).

Since pronunciation influence accuracy and comprehension, it is regarded as a fundamental skill, which students should acquire (Lambacher, 1996). Pronunciation has a considerable influence on the effectiveness of communication; the accuracy of pronunciation ascertains the effectiveness of spoken messages (Singhathin & Wongsaphan, 2021). Pronunciation is the most important and the most difficult part for non-native English speakers (NNES). However, pronunciation instruction has not been included in course outlines per se, and in the majority of English classrooms in Thailand, pronunciation has mostly been overlooked in the teaching and learning process. Insufficient time is allotted for pronunciation given that in most English classrooms in Thailand, grammar is the focus.

Although research studies on speaking skills and pronunciation are prevalent, research on the efficiency of English pronunciation instruction, particularly with Thai L2 learners within a CAPT system context, has somehow been investigated less frequently (Iadkert, 2014). The available CAPT tools are limited in their feedback variety, focus on native-like pronunciation, and have an unreliable automated scoring system. Therefore, this study aimed at developing an instructional pronunciation package that could be suitable for NNSE university students (non-English majors) in higher education, who are at the Beginner and Intermediate levels of English communication. In this package, a CAPT system is employed to assess the efficiency of pronunciation instruction, and interventions are conducted within the CAPT system.

**Research Objectives**

The study aimed at accomplishing the following:

1) investigating the overall efficiency of the pronunciation instructional package,

2) comparing the students' achievement in pronunciation before and after the intervention, and

3) evaluating the students' levels of satisfaction with the instructional English pronunciation package in a CAPT system.

**Research Methodology**

**Participants**

The population of the study was university students (non-English majors) with age ranges of 19-22 years, who were studying Business English in the Academic Year of 2021. Of the students, 425 were from K1 University and 149 were from K2 University. The sample consisted of 30 Thai university students studying English as a Foreign Language (EFL), who were selected by simple random sampling as follows: seventeen were from K1 University (56.67%) and thirteen were from K2 University (43.33%), with the total number consisting of five males (16.67%) and twenty-five females (83.33%). Of them, 14 (46.67%) had finished studying English Pronunciation, 10 (33.33%) had completed Listening and Speaking, and 6 (20%) had studied Public Speaking before participating in this research. The grade results for the 30 students in the previously mentioned courses, which had all been related to communication, were as follows: A (40%), B+ (16.67%), B (20%), C+ (16.67%), C (3.33%), and D+ (3.33%).

**Instruments**

**1. The Application and the Tool**

***Reading Progress (RP)*** is an Automatic Speech Recognition (ASR) that provides separate pronunciation assessment (formative assessment) and feedback on accuracy. RP is used as the principal instrument that students can utilize to read the assigned texts aloud and to make an audio-visual recording of themselves (see Figure 1). ***Immersive Reader (IR)*** is a native-like voice tool that is available on MS Teams and allows students the opportunity to participate in enhanced reading instruction by listening to the voices of English native speakers and by reading along with the given texts so that they can imitate the sounds (see Figure 2). It implements techniques that can improve the students' reading (Jarke et al., 2020), and is used as an accompanying feature with RP. The application and tool were evaluated by three experts, who were EFL teachers. The Content Validity Index (CVI) was 0.70, which was considered to meet the criteria (0.50 to 1.00), and which indicated that the application and the tool were both appropriate for use.

**2. The pronunciation performance test**

A pre-test and a post-test pronunciation performance assessment were used to investigate the participants' pronunciation abilities. A list of 30 words, which contained /-s/ and consonant sounds appearing in one-syllable and two-syllable words that were either in the initial or final position, was compiled as a reading aloud assessment text based upon the most frequently mispronounced by learners. The participants were asked to read those words aloud and to make an audio-visual recording only once via RP without employing the IR at this stage. The participants' levels of English pronunciation performance were evaluated by three raters (one native English speaker and two Thai instructors, who were all specializing in applied linguistics and who were teaching EFL courses). The pronunciation performance test was evaluated, with a rubric score rating 1-5 (1 = serious pronunciation problems, ..., 5 = very clear and easy to understand), by three experts, all of whom were EFL teachers. The CVI was determined to be 1.00, which indicated that the test was suitable to use.

After that, the pronunciation performance test was conducted with a pilot group of 20 students to assess its reliability. The Cronbach's alpha coefficient of the test was 0.86, which verified that the test had achieved an acceptable level of confidence.

**3. The student questionnaire to evaluate the learners' levels of satisfaction with the instructional pronunciation package in a CAPT system** A 5-point Likert scale was designed to be used after the post-test. Its goal was to assess the students' levels of satisfaction with regard to their use of the pronunciation instruction in the CAPT application. A group of three experts evaluated the questionnaire. The CVI was 1.00. Next, the questionnaire was evaluated with a pilot group of 20 students, who had similar characteristics to those in the sample group, to determine reliability. The Cronbach's alpha coefficient of the test was 0.94, which affirmed the questionnaire's acceptable level of confidence.

**4. Reading texts of the pronunciation practices** The exercise plan of the pronunciation practices was prepared and divided into three steps:

**4.1 Content presentation**. Before the actual training began, the application and the tool were demonstrated by describing to the participants how to pronounce the English words along with the IR. The participants were informed that they had the ability to select the gender of the voice of the English native speakers and the speed at which they could listen.

**4.2 Pronunciation exercises**. By selecting twelve business-related reading texts with an average length of 250 words (e.g., company profile, product and service characteristics, marketing, competition, strengths and weaknesses, research and development, unethical business practices, and corporate social responsibility), the pronunciation exercises were formulated. These texts were assigned and uploaded 3 times a week for a month and were generally customized to match the students' reading level (see Figure 1). At the participants' convenience, the pronunciation practices were performed unlimitedly by reading aloud along with the IR (see Figure 2). When the participants felt prepared, they created an audio-visual record of their pronunciation, which ran approximately 30 minutes per practice.

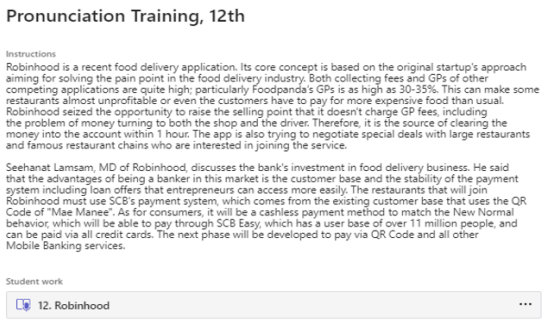


Figure 1

RP: The student's view of the reading text that had been recorded

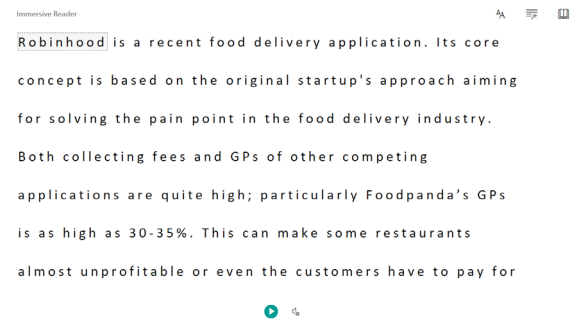


Figure 2

IR: The voice of a native English speaker

**4.3 Score profiles**. At the end of each read aloud assessment text, score profiles were automatically kept and were incorporated with the advice from the human instructor in a weekly one-on-one meeting in Teams, which lasted 10-minutes.

In a comfortable setting, each of the participants was able to read at his/her own pace and to record his/her voice an unlimited number of times, which helped the students to further develop their reading skills at their own individual pace. Given that the app had been streamlined by integrating with Teams' Education Insights dashboard, teachers were able to use the auto-detect feature to quickly review the student's errors (e.g., mispronunciations, repetitions, phrasing, intonations, and omissions) and to override any inaccuracies that the feature may have highlighted. The pronunciation sensitivity level (i.e., high, medium, or low) could be set as demonstrated in Figure 3 to account for different speech patterns and accents. In addition, the number of correct words per minute, the mispronunciations, and the percentages of the accuracy rate were calculated. The accuracy rate of an individual participant could be compared with other class members in the 12-session intervention is illustrated in Figure 4.

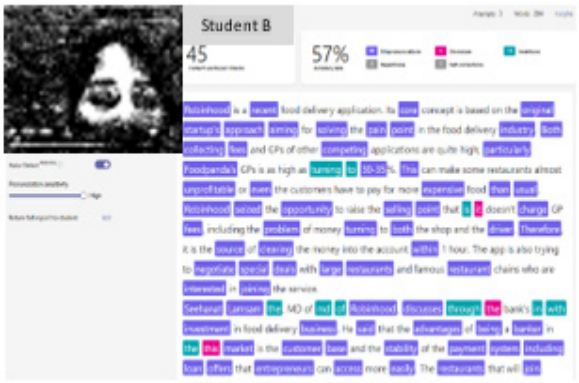


**Figure 3**

RP: The teacher's view (The student's audio-visual clip was recorded with the following: the number of correct words per minute, the accuracy rate, the mispronunciations, the omissions, the insertions, the repetitions, and the self-corrections.)

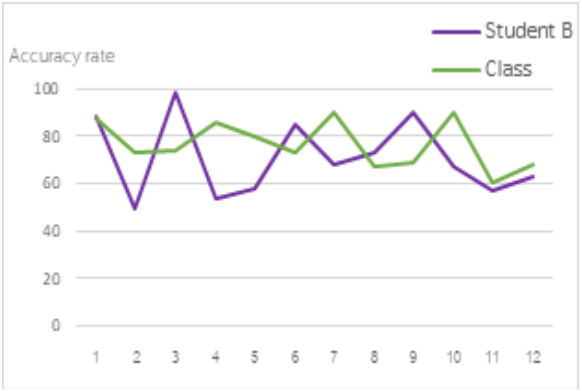


**Figure 4**

The comparison of an individual student's accuracy rate to the other class members during the 12-session intervention

## Data collection

The one group pre-test post-test study was conducted between January and February 2022 through the experimental Teams platform. To examine their pronunciation performance, each participant received a pronunciation performance test before and after the instruction. The duration of the instructional pronunciation package in the CAPT system was 12 interventions over 4 weeks. The efficiency of the instructional pronunciation package was assessed by considering the package's performance, while using the materials throughout the practice exercises and determining the per-

formance after the materials had been used through the end of the unit exercises and the post-test. The tests were conducted before and after the instruction. The study was ethically approved (COE No.65/003) by the university research committee, and consent forms were individually distributed to the volunteers. Regarding the data, obtained from the pronunciation tasks, the tests were manually scored by three raters based upon the pre-determined rubric score. All the data was analyzed using SPSS version 20. In order to determine the effects that the instructional pronunciation package in a CAPT system had had on the participants' pronunciation, the means and standard deviations were ascertained, and the paired t-test was performed.

## Results

*1. To answer the Objective 1, the overall efficiency of the pronunciation instructional package was investigated:*

### *The Instructional Pronunciation Efficiency*

The exercise scores from each practice were designated as $E_1$, and the post-test scores were designated as $E_2$ (Brahmawong, 2013a, 2013b). From Table 1, it was found that the scores from doing the exercises during the training ($E_1$) had an average value of 1,071.87 or 89.32% and the scores from the test after the training ($E_2$) had an average value of 130.20 or 86.80 %. The efficiency index of the process ($E_1$) had been 89.32, and the efficiency index of the product ($E_2$) had been 86.80. The efficiency index ($E_1/E_2$) of the pronunciation exercises, which utilized a pronunciation instructional package in a CAPT system, was 89.32/86.80, which was higher than the specified threshold of 80/80. It can be concluded that the efficiency of the instructional pronunciation package met the specified criteria.

**Table 1**

The instructional pronunciation efficiency

| Efficiency | Full Score | Total Score | Mean | SD | % |
|---|---|---|---|---|---|
| $E_1$ | 1,200 | 32,156 | 1,071.87 | 5.61 | 89.32 |
| $E_2$ | 150 | 3,906 | 130.20 | 0.25 | 86.80 |
| $E_1/E_2$= 89.32/86.80 | | | | | |

*2. To answer the Objective 2, the students' performance in pronunciation before and after the training was tested and compared:*

### *Scores of Pronunciation Exercises*

In regard to the twelve pronunciation exercises, which were completed by the 30 participants, all of the participants had been able to pass the pronunciation assessments, which were derived from the RP's automatic scoring. With the accuracy rate and the manual scoring by the human instructors, the following scores were given during the training: a) a 'very good' level at 56.67% (scores of 1075-1155, out of 1200), b) a 'good' level at 36.67% (scores of 969-1071), and c) a 'moderate' level at 6.66% (scores of 831). This was in accordance with the pronunciation sensitivity level, which had been set in the RP.

*Participant Improvement from Pre-test to Post-test*

In order to investigate normality, the results of the analysis were presented in Table 2 as follows: by Kolmogorov-Smirnov method, sig (pretest) = sig (posttest) = 0.2; and by Shapiro-Wilk method, sig (pretest) = 0.092 and sig (posttest) = 0.136. At $\alpha$ = 0.05, the significant values of both methods were greater than $\alpha$, concluding that the two data sets had a normal distribution.

**Table 2**

Tests of Normality

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | **Statistic** | **df** | **Sig.** | **Statistic** | **df** | **Sig.** |
| pretest | .119 | 30 | .200* | .940 | 30 | .092 |
| postest | .116 | 30 | .200* | .946 | 30 | .136 |

The results showed statistically significant improvement from the pre-test to the post-test by using the t-test, paired two samples for means. The test scores from the pronunciation tasks indicated that the mean score of the post-tests had been higher than the pre-tests ($p < .01$) as presented in Table 3. The mean of the post-test score (4.34) was higher than that of the pre-test score (3.16) with the t-value of 12.40 at the level of significance of 0.00.

**Table 3**

A comparison of the pre-test and post-test pronunciation performance scores

| | **n** | **Mean (Total score=5)** | **SD** | **df** | **t** | **P-value** |
|---|---|---|---|---|---|---|
| Pre-test | 30 | 3.16 | 0.44 | | | |
| Post-test | 30 | 4.34 | 0.25 | 29 | -12.40 | 0.00** |
| Paired | | -0.98 | 0.52 | | | |

*3. To answer the Objective 3, the students' levels of satisfaction with the instructiona l pronunciation package in a CAPT system were evaluated:*

***Satisfaction toward the pronunciation instructional package in a CAPT system***

The learners' level of satisfaction with the instructional pronunciation package in a CAPT system had been strongly positive ( $\overline{x}$ = 4.58, S.D.= 0.52). For the survey questions, there were 2 main categories: 1) the application itself and 2) the human instructor, who had interacted with the participants. Most participants had been strongly satisfied with the following: 1) the functions of IR, which allowed the learners to listen to the native speaker voices and then to repeat after them ( $\overline{x}$ = 4.74, S.D.= 0.53); 2) the function of the RP that had provided them with the opportunity for unlim-

ited attempts at recording ( $\overline{X}$ = 4.56, S.D.= 0.58); 3) the pronunciation score progress ( $\overline{X}$ = 4.52, S.D.= 0.64); 4) the written feedback provided by the instructor ( $\overline{X}$ = 4.67, S.D. = 0.48); and had been satisfied with support from the instructor during the one-on-one meetings ( $\overline{X}$ = 4.41, S.D.= 0.84).

## Discussion

This study was designed to investigate the efficiency of the RP application and the IR tool in improving English pronunciation, to compare Thai EFL students' pronunciation performance as well as their levels of satisfaction with the use of the application.

.The result of the efficiency analysis indicated that with the use of the application, the instruction packagehad been 89.32/86.80, according to the specified 80/80 criteria, which indicated that the instructional package had been profitable for learners seeking to improve their English pronunciation. The process efficiency proved that the following instructional package and the learning process had been effective: 1) practicing pronunciation via the application, 2) receiving written feedback, and 3) consulting with the teacher during the one-on-one sessions. Before implementation, these were validated by five experts. However, the process efficiency was higher than the product efficiency, by more than 5%, which reflected an imbalance between the process efficiency and the product efficiency (Brahmawong, 2013a, 2013b). It was assumed that the post-test might have been too difficult for the students. Another possible cause for the imbalance could have been the students' test anxiety. The students might have perceived the formative assessment as practice, in which they received supportive feedback from the teacher. The post-test scores were not as high as expected. As a result, $E_1$ was much higher than $E_2$.

The differences in the pronunciation performance between the scores of the pre-test and post-test reached a level of significance at $p < 0.05$. Together with the feedback and the guidance provided by the teacher, the application had been able to enhance the students' pronunciation performance. In this study, the students listened to and imitated the computer-generated voices (text-to-speech). Rogerson-Revell (2021) elaborated that using a computer in pronunciation training allows learners to set their learning pace and learn in a stress-free environment. Theoretically, the findings of this study could be explained through the cognitive psychological perspective. According to Skill Acquisition Theory (DeKeyser, 2015), learning has three main stages: learning explicit/declarative knowledge, conversion of declarative knowledge into procedural knowledge, and automatization of knowledge. In this study, great emphasis was placed on the first and second stages of the theory. The application highlighted the mispronounced words as inputs for the teacher to prepare the effective pronunciation instruction in one-on-one feedback sessions. For example, the students learned the declarative knowledge about voiced and voiceless sounds in English to pronounce the words with -ed ending correctly. Besides, the teacher demonstrated the way to pronounce the words, such as talked, wanted, purchased, etc. to establish procedural knowledge. After the students gained the relevant procedural knowledge for improving pronunciation, the learners transformed the procedural knowledge into automatic pronunciation process through practicing. As a result, the post-test score could reflect the students' skill acquisition.

**Technologies** should provide learners with exposure to varieties of the English language and to several accents in order to instill the

learners with intelligibility (Henrichsen, 2021). However, the generated voices from the applications may be based upon native-speaker corpora. Another feature of the application is the Automatic Speech Recognition (ASR), which can highlight the words that have been incorrectly pronounced. ASR seems to be beneficial for scoring. Pennington and Rogerson-Revell (2019), and Rogerson-Revell (2021) claimed that the human biases and human errors often found in pronunciation testing can be avoided by using ASR. However, with respect to the ASR of the application, it is most likely that it solely relies on a database of speech from native speakers. The ability of ASR to recognize non-native pronunciation is far from being flawlessly reliable (Henrichsen, 2021; Rogerson-Revell, 2021).

Due to this limitation, the teacher was required to revise the highlighted words. After the revision, the learners received individualized written feedback given that many researchers (Agarwal & Chakraborty, 2019; Derwing & Munro, 2015; Hincks, 2015; Levis, 2018) have agreed that individualized feedback is essential for pronunciation training. The highlighted errors, the accuracy rates, and the rates of speed were later displayed to each learner. Meanwhile, the verbal feedback and consultations were provided on a weekly basis in one-on-one sessions so that the teachers could elaborate on what learners should do to correct errors and to improve their pronunciation (Henrichsen, 2021), especially with respect to those errors that had become a fossilized portion of the learners' interlanguage. Finally, the learners had been able to improve their pronunciation after the intervention.

The learners in this study derived great satisfaction from the experience of using RP and IR, and from undertaking the learning process. The application, which is available on MS Teams, has been used for online learning since the onset of the COVID-19 pandemic. For this reason, the learners were familiar with the application's interface, and a demonstration was provided prior to the intervention. The features (e.g., the text-to-speech; the voice recordings; the pronunciation error detection, which was displayed in the one-on-one sessions; and the guidance from the teacher) were found to be useful for the learners in improving pronunciation. The application and learning process allowed the learners to control their own learning. Due to their concerns about their inadequate pronunciation skills, some learners might be anxious about learning in a classroom setting (Horwitz, 2010; Nakazawa, 2012).

The results of this study are consistent with those studies, which had employed both a computer and an instructor in pronunciation training. In 2012, Nakazawa found that the participants had been satisfied with the training, in which the use of a computer program was integrated with the instructor-led training. The computer program used in the study was able to provide feedback to each of the participants. However, the participants insisted that the attention that they had received from the instructor on their individual pronunciation was mandatory, and that they regarded the computer program as a supplementary tool. In addition, Gao and Hanna (2016) compared three different interventions, which used: 1) software instruction, 2) human instruction, and 3) a combination of software and human instruction. The results revealed that the participants, who had been trained by using the combined approach, showed the highest level of pronunciation improvement.

## Conclusions, implications, and limitations

The instructional package included twelve reading texts, which had been generated into the voices of native speakers and which was delivered through a RP application and an IR tool. In addition, both written and verbal feedback on the students' pronunciation improvement was given by the instructor. The efficiency of the instructional package had met the expected criterion (80/80). The statistical comparison of the pre-test and post-test scores, which achieved a significant level ($p < .05$), revealed that there had been an improvement in pronunciation. Furthermore, the participants in this study expressed great satisfaction about their experience of utilizing the instructional package.

The application should not be used as a stand-alone tool for pronunciation training because there are some limitations to the applications, which have been mentioned in the discussion. As supplementary tools, the applications provide abundant opportunities for L2 learners to listen and to imitate the computer-generated voices at their own pace, at the best available time, and without any peer pressure and/or anxiety. Because the applications use text-to-speech technology to generate voices, not only could the texts be used as input for pronunciation practice, but they could also be utilized as extensive reading activities. Before selecting an application for pronunciation training, L2 teachers should check its variety of pronunciation feedback, intelligibility-based ASR, and the reliability of automated scoring system. After learning about the limitations of an application, teachers should play a role to transcend the limitations. For example, an application only summarizes the words that are normally mispronounced, without showing how to pronounce them correctly. Most importantly, teachers should play a supportive role in providing sufficient guidance to correct individual errors.

There were some limitations, which were related to the factors affecting L2 pronunciation, sample size, and the research design. Further research should exclude the interference derived from several factors that can affect pronunciation learning, such as age, motivation, attitudes, and L2 aptitude. The small sample size (30 participants) may have threatened the internal and external validity of this study. Therefore, when generalizations are being made, caution should be exercised. It could be assumed that the pronunciation improvement, which was observed after the training, might have resulted from a combination of computer and human instruction. Should further research aim at entangling effects, researchers should employ an experimental research design utilizing a control group and experimental group. A delayed post-test should be integrated into the research procedure in order to determine how long the training effects may last. It appears that binary feedback is the major weakness of RP and IR. Rather than completely relying on the AI's feedback, teachers could play a role in providing individualized feedback. This study elicited the role of teachers in the planning stage, in which the limitations of the software or the applications were investigated, and the available instructional resources were seamlessly integrated into CAPT.

## References

Agarwal, C., & Chakraborty, P. (2019). A review of tools and techniques for computer aided pronunciation training (CAPT) in English. *Education and Information Technologies*, *24*(6), 3731-3743.

Añorga, A., and Benander, R. (2015). Creating a pronunciation profile of first-Year Spanish students. *Foreign Language Annals. 48*(3), 434–446.

Brahmawong, C. (2013a). Developmental testing of media and instructional package. *Silpakorn Educational Research Journal*, *5*(1), 7-20.

_____. (2013b). The efficiency test of instructional media. *Journal of Silpakorn Educational Research, 5*(1), 7-20.

DeKeyser, R. (2015). Skill Acquisition Theory. In B. VanPatten & J. Williams (Eds.), Theories in second language acquisition: An introduction (pp. 94-112). Taylor & Francis.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research*. Amsterdam: John Benjamins.

Gao, Y., & Hanna, B. E. (2016). Exploring optimal pronunciation teaching: Integrating instructional software into intermediate-level EFL classes in China. *Calico Journal*, *33*(2), 201-230.

Henrichsen, L. E. (2021). An illustrated taxonomy of online CAPT resources. *RELC Journal*, *52*(1),179-188.

Hincks, R. (2015). Technology and learning pronunciation. In M. Reed and J. Levis (eds), *The Handbook of English Pronunciation* (pp. 505–19). Malden, NY: Wiley-Blackwell.

Horwitz, E. K. (2010). Foreign and second language anxiety. *Language Learning, 43*, 154-168.

Iadkert, K. (2014). Development of English pronunciation with phonics. Retrieved March 15, 2022 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1023.4197&rep=rep1&type=pdf

Jarke, H., Broeks, M., Dimova, S., Iakovidou, E., Thompson, G., Ilie, S., & Sutherland, A. (2020). *Evaluation of a Technology-based Intervention for Reading in UK Classroom Settings*. RAND.

Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford University Press.

Lambacher, S. G. (1996). Spectrograph analysis as a tool in developing L2 pronunciation skills. *M. Vaughan-Rees (ed.)*, 32-35.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377.

Levis, J. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge University Press.

Maghrebi, F., Heydarpour, M., & Shalmani, H. (2016). The effect of pronunciation training software on the Iranian EFL learners' pronunciation skills. *Modern Journal of Language Teaching Methods, 6*(6), 260–271.

Martin, I. A. (2020). Pronunciation development and instruction in distance language learning. *Language Learning & Technology, 24*(1), 86–106.

Nakazawa, K. (2012). The effectiveness of focused attention on pronunciation and intonation training in tertiary Japanese language education on learners' confidence: Preliminary report on training workshops and a supplementary computer program. I*nternational Journal of Learning, 18*(4), 181-192.

Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning and Technology, 18*(3), 173–192.

Pennington, M. C., & Rogerson-Revell, P. (2019). *English Pronunciation Teaching and Research: Contemporary Perspectives*. London, England: Palgrave Macmillan.

Rogerson-Revell, P. M. (2021). Computer-assisted pronunciation training (CAPT): Current issues and future directions. *RELC Journal*, *52*(1), 189–205.

Schmidt, R. (2001). Attention. In P. J. Robinson (Ed), *Cognition and second language instruction*. Cambridge University Press.

Singhathin, N., & Wongsaphan, M. (2021). Communicative English learning management with multimedia reading instructional package on reading ability and vocabulary achievement for the primary school students. *Practitioner Research*, *3*, 159-169.