Using Online Applications to Enhance Phonetic Acquisition among Learners of Chinese
(网络应用对汉语学习者语音习得的有效性研究)

Xu, Hongying
(徐红英)
University of Wisconsin-La Crosse
(威斯康星大学拉克罗校区)
hxu@uwlax.edu

Li, Yan
(李艳)
University of Kansas
(堪萨斯大学)
yanli@ku.edu

Abstract: Previous literature shows gaps in effective ways to enhance phonetic acquisition among beginning-level learners of Chinese, including tones, initials, and finals, especially in settings outside of laboratories. The present study explored the effectiveness of integrating self-paced use of online applications in non-lab settings in helping students improve their perception and production of Chinese syllables. Three groups of participants were recruited and they underwent training in different settings: one group received four 15-minute sessions using online applications in class; one group completed the same sessions outside of class; the third group received traditional teacher-led instruction and pen-and-paper Pinyin practice in place of the practices using online applications. The two experiment groups performed similarly: both groups showed better retention of their gains after the training ended than the traditional instruction group. Learners also embraced the use of these online applications as effective learning aids.

摘要：文献显示目前对如何帮助中文初级阶段学生有效习得中文语音的研究，特别是在日常课堂环境下进行的习得研究比较缺乏。本研究探讨了网络应用在日常教学中帮助中文入门级学生习得中文语音系统方面所起的作用，包括帮助学生提高辨识、认读中文音节的准确性。一共有三组学生参加了本次研究：一个实验组在课上进行了四次 15 分钟的使用网络应用的拼音练习；另一个实验组在课下进行了四次 15 分钟的使用网络应用的拼音练习；第三组则在实验组使用网络应用进行拼音练习的时间内，在课堂内采用传统教师带领下的语音训练、在课外进行传统的纸笔拼音听力练习。结果发现实验组的学生相较对照组的学生在训练结束后较好地保持了辨识、认读中文音节的准确性。这些学生也把网络应用视为一种有效的学习辅助工具。

Keywords: Online applications, Chinese syllables, perception, production, learning experiences

关键词：网络应用、中文音节、音节辨识、音节认读、学习体验

1. Introduction

Intelligible speech is essential for effective communication, and good pronunciation is an integral part of that, especially for second-language (L2) learners (Arteaga, 2000; Morin, 2007; Thomson & Derwing, 2015). Good pronunciation also helps the construction of a learner's identity in the target language (Levis, 2007). Instruction and training in pronunciation in second language acquisition is indispensable if learners wish to achieve targeted proficiency. However, in the traditional language classroom setting, instruction on pronunciation is often neglected in favor of the development of other skills such as linguistic competence, writing, and so on. (Isaacs, 2009; Thomson, 2011). Studies have shown that students rarely receive sufficient instruction and feedback on pronunciation from their teachers due to a lack of time and/or appropriate resources (Collins & Munoz, 2016; Neri et al., 2010). As a tonal language, Chinese pronunciation system employs different pitch heights and contours to distinguish meaning and has some consonants and vowels that are challenging to English-speaking learners of Chinese (Chun et al., 2015; Jongman et al., 2006). For beginning adult learners, the first few weeks of exposure to the Chinese language constitutes a critical window in the development of their ability to perceive and produce Chinese sounds because adult production and perception systems demonstrate only a certain degree of plasticity (Wang et al., 1999; Wang et al., 2003).

Pronunciation training for Mandarin Chinese is closely related to Pinyin, a transliteration system using Roman letters to represent the sounds of Mandarin Chinese. Pinyin is currently the most popular tool Chinese language learners use in learning Chinese. Chinese learners not only need Pinyin to learn the Chinese sounds, but can also use it as an aid for reading Chinese texts and to key in Chinese characters on their computers (Shei, 2014). Pinyin enables adult non-native speakers to better grasp the Chinese sound system and learn new vocabulary more quickly while developing spoken language proficiency without being intimidated by Chinese characters (Everson, 2011).

The Pinyin system consists of initials, finals, and tones. The initial represents the consonantal beginning of a syllable. Finals are mainly vowels or vowels with nasal endings that follow the initial in a syllable. There are four tones in Mandarin Chinese, namely the high-level tone (T1), the high rising tone (T2), the dipping tone (T3), and the high falling tone (T4). These four tones are represented by diacritic marks that are placed over the main vowel in each syllable. The Pinyin system seems to be simpler than Chinese characters as a way for English-speaking L2 Chinese learners to learn pronunciation as it employs the same Roman letters used in English. However, learning Pinyin is not as simple as many people believe. In Pinyin, one letter may represent many different phonemes which sound very different to non-Chinese speakers. For example, the letter *e* can either be transcribed as [ê] or as [e]. The same Roman letter may represent similar yet different sounds in Mandarin Chinese and English. For example, the letter c is pronounced [ts^h] in Chinese, whereas it is pronounced [k^h] in English most of the time. Learning Pinyin means essentially learning a new orthographic system: learners need to establish the correct sound-symbol mapping while acquiring the Chinese pronunciation system. Li and Xu (2018) did a survey among instructors of Chinese and found that the majority of them

agreed upon the importance of learning Chinese Pinyin while lamenting the limited instruction time that can be devoted to teaching Pinyin in class. It would seem, then, that Chinese teachers all face this question: How to assist Chinese learners most efficiently in mastering Pinyin given the limited instruction time? Computer-assisted pronunciation training (CAPT) might be an answer. Levis (2007) reviewed major empirical studies on CAPT to discuss how technological tools that have long been used in phonological training can also be applied to teaching. These studies show that CAPT is effective in improving pronunciation accuracy and that CAPT learning transfers to novel contexts and lasts over time. There is also evidence showing that improvements in perception can lead to improvements in production (Bradlow et al., 1997). Levis (2007) pointed out that CAPT can provide individualized instruction outside of class and can also help those instructors who feel unprepared to teach pronunciation due to lack of proper training.

There have been studies on how CAPT can help L2 Chinese learners acquire the Chinese pronunciation system (e.g., Beutner, 2001; Chun et al., 2015; Godfroid et al., 2017; Wang et al., 2003). However, most of the studies focus on tone acquisition and few studies focus on the acquisition of consonants and vowels. This study, aiming to provide a possible solution to the current dilemma in Pinyin instruction and phonetic training, investigated whether CAPT can help L2 Chinese learners establish the mapping between the Chinese pronunciation system and Pinyin, including consonants, vowels, and tones.

This article is organized as follows: Section 2 provides an overview of the research on pronunciation instruction (PI) and the integration of CAPT and Chinese phonological training. Sections 3, 4, 5, and 6 describe the current empirical study, including research questions, methods, results and discussion. Section 7 draws a conclusion based on the experiment and discusses the pedagogical implications of this study. Section 8 considers the limitations of the current study and suggests directions for future research.

2. Research on Pronunciation Instruction (PI) and the Integration of CAPT and Chinese Phonological Training

Most studies on pronunciation instruction (PI) and the integration of CAPT and PI have investigated its effectiveness on learning segmental and suprasegmental features of L2 English. Saito (2012) did a synthesis of 15 quasi-experimental studies that explored the effects of instruction on the development of L2 pronunciation. Ten of the studies focused on L2 English, three on Spanish, one on French and one on Portuguese. None was on L2 Chinese. Lee et al. (2015) did a meta-analysis of PI effects based on 86 studies. They pointed out that 83 out of the 86 studies involved either L1 or L2 English and thus a more diversified sampling in research along this line was called for. Mahdi and Khateeb (2019) did a meta-analysis on the effectiveness of CAPT based on 20 studies. Nineteen out of the 20 studies focused on L2 English. Because the relation between L1 and L2 has an impact on the acquisition of L2 pronunciation (Flege, 1995; Thomson, 2011), the skewed sampling (a majority of the studies in L2 English) in PI and CAPT research could be affecting the generalizability of the findings from these studies to the instruction of pronunciation in other target languages, including L2 Chinese.

Among the studies on the training of students in Chinese pronunciation, there are many that focus on helping students improve either their perception or their production of Chinese tones, and the majority of these studies concern themselves with auditory training (Zhang, 2018). Wang et al. (1999) found that intensive tone training in a lab setting could improve L2 Chinese learners' perception of tones. This kind of training can also facilitate production (Rochet, 1995; Wang et al., 2003). A similar trend appears in CAPT research on the training of L2 Chinese pronunciation. Most studies focus on the perception or production of tones after some CAPT intervention. Beutner (2001) found that using computer-assisted interactive feedback helped L2 Chinese learners improve their tone production. Chun et al. (2015) found that L2 Chinese learners' tone production improved after they compared the visualizations of the tones they produced with those created by native-speakers. Godfroid et al. (2017) explored whether the types of cues, used alone or together, had an impact on the training effects of learners' perception of Chinese tones and they found that dual-cues did not offer an advantage over single-cues and that pitch contours and numbers worked better as cues than colors. However, the studies were all conducted in a lab setting. Can similar but more flexible training be conducted out of the lab to assist learners' acquisition of the Chinese pronunciation system? As Thomson (2011) points out, there is a gap between the findings from experimental studies and actual classroom implementation. More studies in classroom settings with training procedures that can be mimicked more easily in similar situations are clearly needed.

However, some studies (Olsberg et al., 2007) discovered a dependence of tone perception on syllable perception. They found that once the spectral information was removed or filtered, subjects' ability to identify the tones decreased. Sharma et al. (2015) found that native speakers of Chinese had as much trouble identifying finals in non-word syllables as they did identifying tones. Lin and Lin (2010) found that vowel information was available earlier than tonal information in native Chinese speakers' perception of monosyllables. Hu (2009) found that both native speakers of Chinese and L2 Chinese learners performed better in perceiving a whole syllable than in perceiving tones, initials, and finals. There was no significant difference in their perception between tones and finals. Pytlyk (2011) established that orthographic information exerts influence on L2 learners' perception of phonemes and whole syllables. All of the studies indicate that initials and finals should be part of the training process though more research is needed on how to help learners improve their perception of initials and finals.

Research from previous studies in the areas of both PI and CAPT has shown that production is closely linked to perception (Baker & Trofimovich, 2006; Flege, 1995; Thomson, 2011). As Thomson (2011) puts it, "in most cases, pronunciation inaccuracies betray underlying perceptual inaccuracies" (p.744), therefore, "improvement in perception should allow learners to more effectively monitor their own productions" (p.749). Previous studies have shown that training in perception can result in an improvement in production (Bradlow et al., 1997; Lambacher et al., 2005; Thomson, 2011; Wang et al., 2003). Bradlow et al. (1997) found that, after receiving perception training in distinguishing between the two English consonants /l/ and /r/, L1 Japanese learners of L2 English improved their pronunciation of these two consonants. Wang et al. (2003) noted an 18% improvement in the pronunciation of Mandarin tones among L1 English L2 Chinese

learners after perceptual training. Thomson (2011) found that, after receiving computer-based high variability phonetic training (HVPT) in the pronunciation of 10 English vowels, L1 Chinese L2 English learners significantly improved their pronunciation of these vowels as measured by intelligibility. However, these studies either focused on L2 English consonants or vowels, or tone production among L2 Chinese learners. Few studies have looked into whether computer-assisted perception training might help L2 Chinese learners improve both their perception and their production of Chinese consonants, vowels and tones.

The present study fills the gaps by investigating whether CAPT used in a non-lab setting can help L2 Chinese learners establish the mapping between the Chinese pronunciation system, including tones, initials and finals, and Pinyin, and whether perception training for learners can help their production in these aspects.

3. Purpose of the Current Study

Our study addresses some of the gaps identified in the discussion above and explores the way that integrating computer-assisted training programs – online applications in this study – into the language instruction curriculum can help students improve their perception and production of Chinese syllables, including initials, finals) and tones. Specifically, this study addresses the following questions:

1) Can online applications help students improve their overall perception of Chinese initials, finals and tones?
2) Can online applications help students improve their production of Chinese initials, finals and tones?
3) Does the context of the training (i.e., in class or outside of class) have an impact on the training results?
4) How well are the online training applications accepted by learners of Chinese?

4. Methods

The study adopted a quasi-experimental design, including a pretest, the training, a post-test, and a delayed post-test. A delayed post-test was included to see if any training effects were retained after the training period.

4.1 Participants

Twenty-six English-speaking learners of Chinese from two universities in the American Midwest participated in the study. They were all beginning-level Chinese learners who had no experience learning Chinese or any tonal language before the training started. Three intact classes were used as convenient samples. The classes were put into three settings: Group 1 received the treatment (i.e., use of online applications) in class, Group 2 received the treatment outside of class, and Group 3 did not receive treatment either in class or outside of class; instead, they received traditional instructor-led training

in class (i.e., no online applications) and did traditional pen-and-paper Pinyin practice outside of class to make sure their Pinyin practice time was comparable to that of the two experimental groups. Group 1 and Group 3 were from one university and Group 2 was from another university. The groups were assigned in this manner so that the differences observed, if any, would more likely arise from training rather than possible differences in the student population, although these two universities are both located in Midwest and have a similar student composition. The pre-test showed that there was no significant differences among the groups. The study started with 33 qualified participants, however, seven participants either did not complete one of the tests or missed a training section and were thus excluded from our analysis. Table 1 summarizes the distribution of participants and their demographic information. None of the participants had any history of hearing, speech, or language difficulties.

Table 1 Demographic summary of participants

Group	*No.*	*Gender*	*Age*
Group 1: Training in class	12	M=6; F=6	20
Group 2: Training outside of class	5	M=3; F=2	19
Group 3: Traditional instruction	9	M=4; F=5	19

4.2 Materials

Two online applications were selected as training material: one is the Pinyin learning application (Pinyin Practice hereafter) embedded in the website http://www.pinyinpractice.com; the other is Pinyin Tutor (https://sla.talkbank.org/pinyin/), a website developed by a research group at Carnegie Mellon University. Pinyin Practice provides practice in different categories: tones, initials, and finals. There are also self-administered quizzes that ask students to type in the Pinyin of the sound they hear, including tones. For the practice part, learners listen to the syllables, with a choice of screen display – either characters or Pinyin, or both, or none – and then they either pick or enter the target part (tone, initial, or final). Instant feedback is given (correct or incorrect) after each attempt. Correct answers are displayed after a certain number of attempts that can be pre-set by the learner in the self-administered quiz mode, but not in the practice mode. A running tally of both successful and unsuccessful attempts is displayed throughout the process.

Pinyin Tutor provides students with a similar platform on which to practice their perception of Chinese pronunciation and to match it with Pinyin orthography. Learners listen to a target syllable or word and then type the Pinyin into a text box to indicate what they hear. If what is entered matches the sound that was produced, positive feedback (congratulations) will be given and the next target item will be presented. If what is entered is not correct, the Pinyin Tutor gives feedback on which component of the entered syllable is incorrect and lets the learner try again. Learners can also click on “Listen to Your Attempt” button to hear the sound that matches the Pinyin they have entered so that they can compare the two syllables and notice the difference. In addition, Pinyin Tutor automatically collects the incorrectly perceived syllables and keeps them through subsequent practice rounds until all the syllables are correctly entered in Pinyin. These two online applications were chosen because they are both free and can be used on any digital

device. More importantly, they both provide instant feedback, which is considered one of the most beneficial and important features for CAPT (Lee et al., 2015; Levis, 2007; Thomson, 2011.)

4.3 Measurement of Perception and Production

Participants completed one perception task and one production task on each pre-test, post-test and delayed post-test. The perception task (Task 1) had three sections, with 22 tokens in each section. In Section 1, participants listened to a syllable three times and then wrote down the initial and tone they heard as the final of the syllable was provided. Section 2 followed the same design but elicited responses only to finals and tones. Section 3 elicited responses to the whole syllable. The syllables were not selected from any textbooks. Eleven initials and 11 finals were used in the perception task. The initials consisted of two subsets: one that was easy to identify and one that was more challenging to identify as determined by findings from previous research (Hu, 2009; Pytlyk, 2011). There were two categories of finals: simple vowels, and compound vowels. The initials and the finals were then put together so that a) each initial and each final appeared four times, and b) each syllable was a meaningful syllable in Chinese (See Appendix A for a complete list of the syllables used in Task 1). The syllables used in Sections 1 and 2 were the same, though given in a different order. The tokens in Section 3 were different from those in Sections 1 and 2.

The production task (Task 2) consisted of the 44 syllables that were used in Task 1 (see Appendix B). Participants were asked to read the syllables aloud as accurately as they could. They recorded their readings and then submitted the files to a learning management system.

4.4 Procedure

Because this study was conducted at two universities instead of one, extra effort was made to keep the research context as consistent as possible. During the period in which this study was conducted, instructors of these three participating classes strictly followed the same class schedule and used the same set of instructional materials including the PowerPoint presentation on Pinyin, classroom practice materials, and assignments. These three groups differed only in the way they did Pinyin practice (using online apps vs. not using them; using online apps in class vs. using them outside of class), which is the focus of this study. In this way, except for the training methods, the input and feedback participants received was as similar as possible under the circumstances. Instructors briefly introduced Pinyin and the Chinese pronunciation system before the training started. All three groups of participants signed consent forms and then took the pre-test. In the pre-test, Task 1, the perception test, was completed in class and Task 2, the production test, was assigned as homework. The training period included four 15-minute sessions implemented over two weeks. During the training period, participants received either instructor-led in-class Pinyin training or training using online applications in class.

Group 1 participated in the training sessions using the online applications in class: they used Pinyin Practice for the first two sessions and Pinyin Tutor for the remaining two. After class, they completed worksheets on Pinyin. These worksheets provided practice on tones, initials and finals and they were designed to be completed in 15 minutes. Group 2 received traditional instructor-led Pinyin practice in class, during which the instructor listened to students' pronunciation and gave corrective feedback when needed. After watching tutorial videos that showed them how to use the online applications, participants in Group 2 completed the training sessions as after-class assignments. They were asked to practice Pinyin on Pinyin Practice or Pinyin Tutor for 15 minutes for 4 days and report the time they spent and any issues they encountered. Group 3 worked as the control group. They did not use either application, neither in class nor after class. Instead, they received traditional instructor-led practices in class, similar to what Group 2 received, and completed the same worksheets as Group 1 did after class. This design was created to ensure that the Pinyin practice each group received and completed was comparable.

Table 2 Training design and training materials

Groups	*In-class practices*	*Outside-class practices*
Group 1	On-line applications	Pinyin worksheets
Group 2	Instructor-led Pinyin practice	On-line applications
Group 3	Instructor-led Pinyin practice	Pinyin worksheets

All of the groups took a post-test right at the end of the training period and then took a delayed post-test four weeks later. Both the post-test and the delayed post-test consisted of a perception task and a production task. The tokens used in the post-test and the delayed post-test were the same as the ones used in the pre-test, but in a different order.

After the delayed post-test, a survey was given to the students in Groups 1 and 2 to collect feedback on the use of the online applications. The survey had multiple choice questions, Linkert-scaled questions, and open-ended questions. Please see Appendix C for the whole survey.

4.5 Data Coding

The authors graded Task 1, the perception task, manually. In this task, students heard a sound and then provided a missing initial, final and/or tone. The participants received 1 point for each correct initial, final and tone. Accuracy rates were calculated for a) each category: initials, finals, and tones; b) each participant, and c) each specific initial, final, and tone. The two authors graded all the answers separately and then compared the results: the two sets had been graded exactly the same way. Thus, the inter-rate reliability for this part was 1.

For Task 2, the production task, three raters in total were involved in the rating process. Two raters listened to the syllables and graded the initial, final, and tone of each syllable on a 5-Linket scale with 5 being native-like and 1 being totally unintelligible. Each rater independently rated all 26 participants on all 44 syllables recorded in the pre-test, post-test and delayed post-test. If there was only a one-point difference between the two raters, the average of the two scores was used. If the difference was more than one point,

the third rater was enlisted to listen to those syllables and rate them and the average of the three scores was used.

All of the data were first recorded in an Excel file and then exported to SPSS for statistical analysis. The student survey was distributed in hard copy and responses were entered in an Excel file and manually analyzed question by question. For the perception and production data, mixed measures ANOVAs were conducted to compare the means. The small number and uneven distribution of participants in each group reduced the power to detect significant differences, but the assumptions behind the decision to use ANOVAs were satisfied. A Quantile-quantile plot of residues was created and did not show any significant departure from normality. In other words, no group sample deviated significantly from normal distribution values. Adjustments were made when the Test of Equality of Covariance was significant.

5. Results

The results are reported based on the research questions listed in the previous section. The scores from the overall perception and from specific components, namely initials, finals, and tones, are reported first, followed by the overall production scores and specific scores from initials, finals and tones.

5.1 Results of the Perception Tests

The overall performance results of the three groups across the three perception tests are summarized in Figure 1 and Table 3. Group 1 and Group 2 show a similar trend: their perception accuracy increased over time: from pre-test, to post-test, and to delayed post-test. Meanwhile, the standard deviation declined along these tests, which means the performance within each group became more consistent. However, the same trend was not found among the students in Group 3, whose perception accuracy showed an increase on the post-test but a decrease on the delayed post-test. Likewise, the standard deviation fluctuated among the tests, which indicates that the variation within the group remained large.

Table 3 Perception accuracy rates among the groups across the three tests

Groups	Pre-test		Post-test		Delayed Post-test	
	M	SD	M	SD	M	SD
Group 1 (Training in class) n=12	60.2	17.9	81.2	13.2	83.4	9.8
Group 2 (Training outside class) n=5	55.1	25.5	79.6	6.4	84.8	6.0
Group 3 (Traditional instruction) n=9	53.8	17.2	79.9	11.5	77.2	14.4

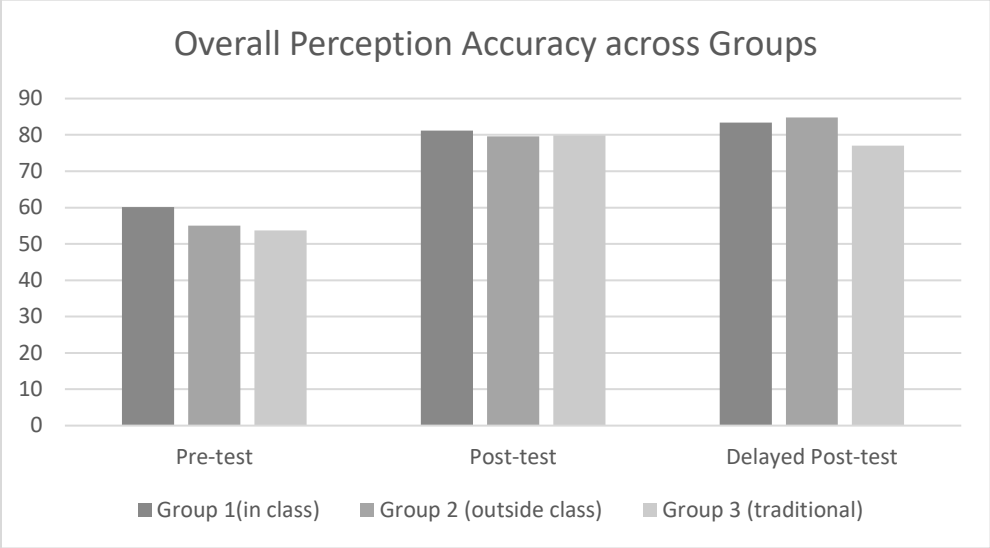

Figure 1 Perception accuracy among the groups across the tests

In order to see whether the differences among the groups within each test and across the tests were statistically significant, a mixed measures ANOVA was conducted with the within-group factor having three levels: pre-test, post-test and delayed post-test and with group as the between-subject factor. When the mixed measures ANOVA was first run, Box's Test of Equality of Covariance was significant ($p = .011$). After removing two outliers from Group 3 and one outlier from Group 1, the equality of covariance was not significant ($p = .097$). The assumption was thus satisfied. Since Mauchly's Test of Sphericity was also significant ($p = .002$), the adjusted values from Greenhouse-Geisser corrected analysis are reported here. There was a significant main effect of time: $F(1.35, 19) = 64.80$, $p < .001$. The results of pair-wise comparisons show that the accuracy rates on the post-test and the delayed post-test were significantly better than those on the pre-test. There was no significant difference between the delayed post-test and the post-test. However, there was no interaction effect between time and group: $F(2.70, 40) = .57$, $p = .69$. The results indicate that all of the groups did better after the training, and their gains were largely retained in the long term. However, the improvement across the tests did not differ significantly among the groups.

The perception accuracy rate for each category – initials, finals, and tones – was also calculated by sections in the perception test: Section 1 asked students to provide the missing initials and tones, Section 2 asked students to provide missing finals and tones, and Section 3 asked students to provide the whole syllable. The purpose of this approach was to determine whether there was any difference in the accuracy rate under different settings: either with part of the syllable provided or without. Therefore, the accuracy of initials and finals was calculated in two settings: either with finals or initials provided or without. In other words, for initials, Setting 1 means finals were provided and Setting 2 means nothing was provided, and, for finals, Setting 1 means initials were provided and Setting 2 means nothing was provided. The accuracy of tones was calculated in three settings: Setting 1 where an initial was provided, Setting 2 where a final was provided, and

Setting 3 where nothing was provided. The results of participants' perceptions of initials, finals and tones in different settings are summarized in Table 4, 5, and 6 respectively.

Table 4 Initials accuracy rates among the groups across the three tests

Groups	Pre-test		Post-test		Delayed Post-test	
	Setting1 M (SD)	Setting 2 M (SD)	Setting1 M (SD)	Setting 2 M (SD)	Setting 1 M (SD)	Setting 2 M (SD)
Group 1 (Training in class) n=12	57.6 (13.2)	60.2 (9.9)	78.0 (17.2)	80.7 (11.2)	80.7 (11.2)	79.9 (11.2)
Group 2 (Training outside class) n=5	68.2 (11.1)	63.6 (19.5)	73.6 (10.9)	75.5 (11.9)	79.1 (12.7)	81.8 (5.6)
Group 3 (Traditional instruction) n=9	48.0 (20.4)	55.1 (11.7)	78.8 (13.3)	80.3 (11.8)	74.2 (12.7)	76.8 (14.1)

Table 5 Finals accuracy rates among the groups across the three tests

Groups	Pre-test		Post-test		Delayed Post-test	
	Setting1 M (SD)	Setting 2 M (SD)	Setting1 M (SD)	Setting 2 M (SD)	Setting 1 M (SD)	Setting 2 M (SD)
Group 1 (Training in class) n=12	64.8 (20.2)	46.2 (22.3)	77.7 (22.3)	72.7 (16.6)	85.2 (16.5)	75.8 (18.1)
Group 2 (Training outside class) n=5	60.9 (13.5)	48.2 (22.4)	70.0 (6.9)	68.2 (12.4)	84.5 (8.3)	68.2 (14.0)
Group 3 (Traditional instruction) n=9	65.7 (17.5)	40.4 (14.5)	81.3 (14.5)	72.7 (16.9)	85.9 (12.1)	70.7 (18.6)

Table 6 Tones accuracy rates among the groups across the three tests

Groups	Pre-test			Post-test			Delayed Post-test		
	Setting 1 M (SD)	Setting 2 M (SD)	Setting 3 M (SD)	Setting 1 M (SD)	Setting 2 M (SD)	Setting 3 M (SD)	Setting 1 M (SD)	Setting 2 M (SD)	Setting 3 M (SD)
Group 1 (Training in class) n=12	61.0 (29.1)	65.2 (27.0)	66.3 (22.6)	88.6 (11.5)	85.6 (12.2)	85.2 (12.6)	85.6 (12.5)	89.0 (7.6)	87.5 (11.3)
Group 2 (Training	45.4 (48.1)	45.4 (44.9)	53.6 (42.9)	90.0 (8.1)	90.9 (9.6)	89.1 (10.0)	91.8 (9.9)	94.5 (3.8)	93.6 (6.9)

outside class) n=5									
Group 3 (Traditional instruction) n=9	55.6 (22.8)	53.5 (31.1)	58.1 (28.2)	82.3 (18.4)	83.8 (17.2)	79.8 (12.9)	79.3 (17.9)	77.3 (21.9)	75.3 (21.9)

The breakdown into initials, finals and tones followed a similar trend. Each group showed an improvement in their accuracy in perceiving the segments and tones immediately after the training period. Mixed measures ANOVAs were conducted with two within-group factors: time, and setting, and one between-group factor: group (training method). The results also showed a similar trend: the main effect of time was significant with initial ($F(2, 22) = 31.37$, $p < .001$), finals ($F(1, 23) = 44.55$, $p < .001$), and tones ($F(2,22) = 21.06$, $p < .001$). There were no significant differences between the groups in perceiving initials, finals, and tones. These results indicate that participants from all groups improved their perception significantly over time, however, this improvement did not differ significantly among the groups. However, the numerical values on the delayed post-tests show that the two groups that used applications either retained or improved their accuracy, whereas the traditional instruction group showed some decrease in their perception of initials, finals (in the setting of "initials provided"), and tones.

All of the groups showed similar accuracy in perceiving initials and tones in different settings, i.e., when either part of a syllable was provided or was not. However, the accuracy in perceiving the finals did show a significant interaction between time and setting ($F(2, 22) = 8.95$, $p = .001$.): participants from all groups increased their accuracy in perceiving the finals over time, but, although the accuracy rates were still higher with Setting 1 than with Setting 2, this increase was significantly greater in the setting in which initials were not provided (Setting 2) than in the setting in which they were provided (Setting 1). And, the performance of the two online application groups on the delayed post-test, in particular, showed that they had retained their improvement.

In summary, results from both the overall perception accuracy and the breakdown into categories of initials, finals, and tones show that there was significant improvement after the training, whether it was received through online application or traditional instruction. However, as indicated by the results of the delayed post-test, the groups that used online applications as the training method not only retained their improvement but showed further progress. Although this difference was not significant, the trend was observed across all the sub-categories. As for the settings in which the initials and tones were elicited, there was no significant difference between them. However, the setting did have a significant impact on the perception of finals: the accuracy was significantly higher when initials were provided than when they were not. However, the online applications helped them improve more in the setting without initials than in the one with initials provided.

5.2 Results in the Production Tests

In the three production tests, participants read aloud a list of 44 Chinese syllables and recorded their readings. Each syllable was rated separately by initials, finals, and tones on a 5-point Linkert scale. The overall production scores of each group across the pre-test, post-test, and delayed post-test are reported first followed by performance scores which are broken down into specific categories: initials, finals, and tones.

The results of the overall performance of the three groups across the three production tests are summarized in Figure 2 and Table 7. The raw scores were based on a 5-point Linkert-scale. They were transformed into percentages to make them comparable with the perception accuracy rates. The results indicate that all of the groups improved their overall production of Chinese syllables over time. The traditional instruction group showed a greater increase on the post-test than the two groups that used online applications. However, the two groups that used online applications showed a greater improvement than the traditional instruction group on the delayed post-test.

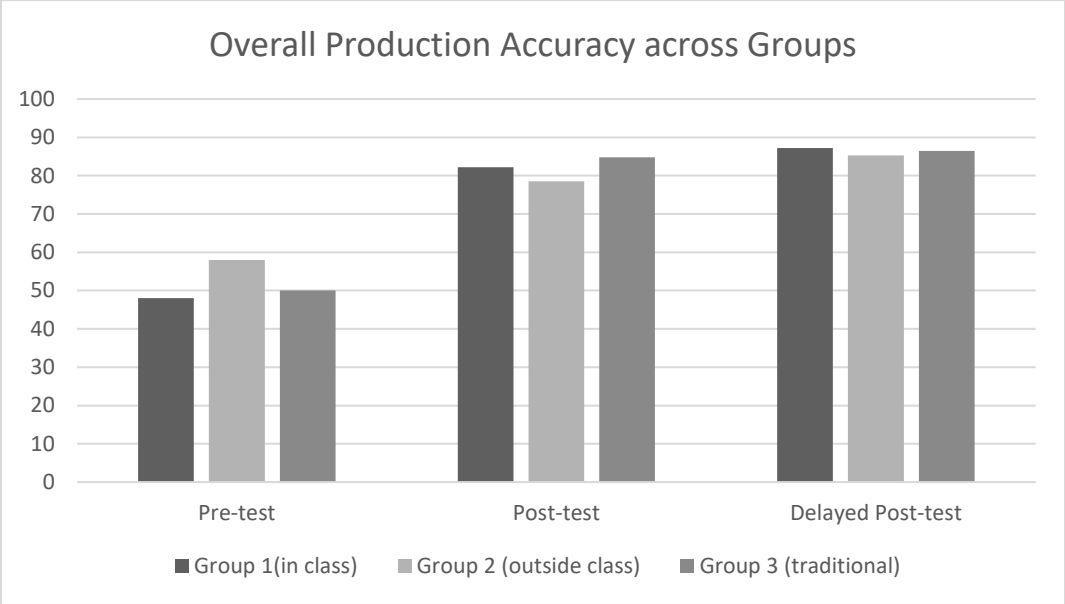

Figure 2 Production scores (%) among the groups across the tests

Table 7 Production scores (%) among the groups across the three tests

Groups	Pre-test		Post-test		Delayed Post-test	
	M	SD	M	SD	M	SD
Group 1 (Training in class) n=12	48.0	17.4	82.2	11.6	87.2	6.9
Group 2 (Training outside class) n=5	58.0	14.5	78.5	9.3	85.3	6.4
Group 3 (Traditional instruction) n=9	50.0	17.1	84.8	10.9	86.5	7.7

In order to see whether the differences among the groups within each test and across the tests were statistically significant, a mixed measures ANOVA was conducted with the within-group factor having three levels: pre-test, post-test and delayed post-test and with group as the between-subject factor. Since Mauchly's Test of Sphericity was significant ($p = .003$), adjusted values from Greenhouse-Geisser corrected analysis are reported here. There was a significant main effect of time: $F(1.42, 22) = 151.16$, $p < .001$. The results of pair-wise comparisons show that the production scores from the post-test and the delayed post-test were significantly better than the pre-test ($p < .001$). The scores on the delayed post-test were significantly better than those on the post-test ($p = .012$). However, there was no interaction effect between time and group: $F(2.85, 44) = 2.19$, $p = .11$. The results indicate that all of the groups did better after the training, and their improvement continued after the training. However, the improvement across the tests did not differ significantly among the groups.

The production scores from each category – initials, finals, and tones – were also calculated. Since the results from each category were different, they are reported here one after another. Each initial, final, and tone from each syllable was rated on a 5-point Linkert scare. An average score was calculated for each participant regarding their pronunciation of the initials, finals, and tones on each test. The purpose of doing this was to gather more details about their production and their improvement after their training in order to determine whether their progress was even across these categories. The results of participants' production of initials are summarized in Table 8.

Table 8 Scores on production of initials among the groups across the three tests

Groups	Pre-test		Post-test		Delayed Post-test	
	M	SD	M	SD	M	SD
Group 1 (Training in class) n=12	49.2	14.7	84.9	9.9	85.9	8.4
Group 2 (Training outside class) n=5	63.7	12.1	80.1	9.7	84.6	4.4
Group 3 (Traditional instruction) n=9	48.8	17.0	87.9	8.0	85.6	7.1

All of the groups showed an increase in their production scores on the post-test. The groups that used online applications (Group 1 and Group 2) showed less improvement on the post-test given immediately after the training than did the group who received traditional instruction (Group 3). However, both online application groups showed some improvement on the delayed post-test, whereas the traditional group showed some decrease. A mixed measures ANOVA was conducted with one within-group factor: time (three levels: pre-test, post-test, and delayed post-test) and with group as the between-group factor. There was a significant main effect of time: $F(2, 22) = 74.71$, $p < .001$. There was no significant main effect of group ($F(2, 23) = .164$, $p = .85$. However, there was a significant interaction effect: $F(4, 44) = 2.77$, $p = .038$. These results indicate that participants from all groups increased their scores in pronouncing the initials over time, but this increase was significantly different among groups over time, as discussed above.

Table 9 Scores on production of finals among the groups across the three tests

Groups	Pre-test		Post-test		Delayed Post-test	
	M	SD	M	SD	M	SD
Group 1 (Training in class) n=12	47.8	16.1	78.9	15.4	84.5	8.6
Group 2 (Training outside class) n=5	52.4	12.2	72.9	8.1	81.8	7.3
Group 3 (Traditional instruction) n=9	50.3	18.9	80.0	12.2	84.9	9.1

The results of participants' production of finals are summarized in Table 9. All of the groups showed improvement in their pronunciation scores on both the post-test and the delayed post-test. The group that used online applications in class (Group 1) and the traditional instruction group (Group 3) showed equal improvement in their scores, whereas the group that used online applications outside of class (Group 2) showed less improvement. The reverse trend was observed on the delayed post-test: the groups that used online applications, especially the one that used them outside of class, showed more improvement. A mixed measures ANOVA was conducted with one within-group factor: time (three levels: pre-test, post-test, and delayed post-test) and with group as the between-group factor. There was a significant main effect of time: $F(2, 22) = 72.41$, $p < .001$. There was no significant main effect of group: $F(2, 23) = .09$, $p = .911$. There was no significant interaction effect between time and group: $F(4, 44) = .77$, $p = .552$. These indicate that participants from all groups showed a similar improvement pattern on their finals pronunciation over time.

Table 10 Scores on production of tones among the groups across the three tests

Groups	Pre-test		Post-test		Delayed Post-test	
	M	SD	M	SD	M	SD
Group 1 (Training in class) n=12	47.1	24.9	82.7	16.4	91.2	10.6
Group 2 (Training outside class) n=5	58.1	20.8	82.7	13.1	89.5	9.2
Group 3 (Traditional instruction) n=9	51.0	21.0	86.5	18.6	88.9	13.7

The results of participants' production of tones are summarized in Table 10. Similar to the scores on finals, the scores on tone production also increased among all groups on both the post-test and the delayed post-test. Both the group that used online applications in class and the traditional instruction group showed equal improvement right after the training, whereas the group that used online applications outside of class showed less improvement. However, both groups that used online applications showed greater improvement on the delayed post-test than the traditional instruction group did. A mixed measures ANOVA was conducted with time as the within-group factors at three levels, pre-test, post-test, and delayed post-test, and with group as the between-group factor. There was a significant main effect of time: $F(2,22) = 51.62$, $p < .001$. There was no significant group main effect or interaction effect between time and groups. These results indicate that

participants from all groups improved their production of tones significantly over time; however, this improvement did not differ significantly among the groups.

In summary, both the overall production accuracy and the breakdown of initials, finals, and tones indicate a significant improvement after training, whether it was through online applications or traditional instruction. However, compared with the traditional instruction group, the groups that were trained using online applications showed a more robust retention trend as indicated by their performance on the delayed post-test. Although this difference was not significant, the trend was observed across all the sub-categories.

5.3 Results of the Two Groups that Used Online Applications in Different Contexts

The results from the previous two sections include the two groups that used online applications but used them in different contexts. One group (Group 1) of participants used them in class under the instructor's guidance. The other group (Group 2) used them outside of class as assignments, and completed a self-study report. The results of their perception and production of Chinese syllables accuracy tests are re-presented in Figures 3 and 4 for readers' convenience.

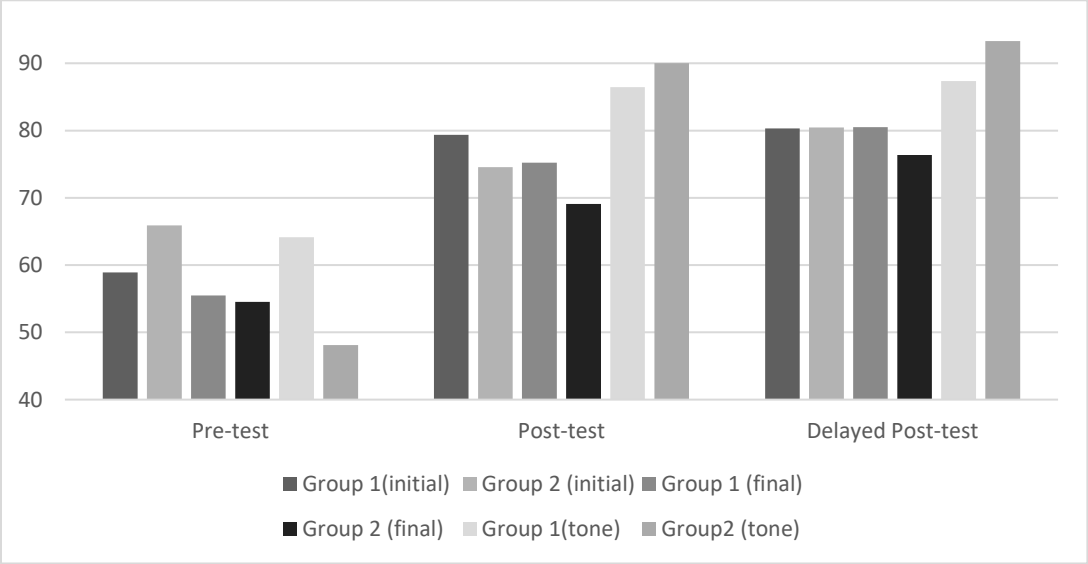

Figure 3 Perception of Chinese syllables between the two groups using online applications

As for perception, the descriptive data show a very similar improvement in both groups' accuracy over time, except for the accuracy improvement on tones exhibited by Group 2, which started at a lower rate and increased more dramatically than that of Group 1. A mixed measures ANOVA was conducted with two within-group factors: time (three levels: pre-test, post-test, and delayed post-test) and category (three levels: initial, final, and tone). There was no significant group main effect ($F(1, 15) = .066$, $p = .801$) or interaction effect among time, category, and group ($F(2.013, 12) = 3.141$, $p = .057$). The two groups showed no significant difference in their improvement in accuracy across the initials, finals, and tones over time.

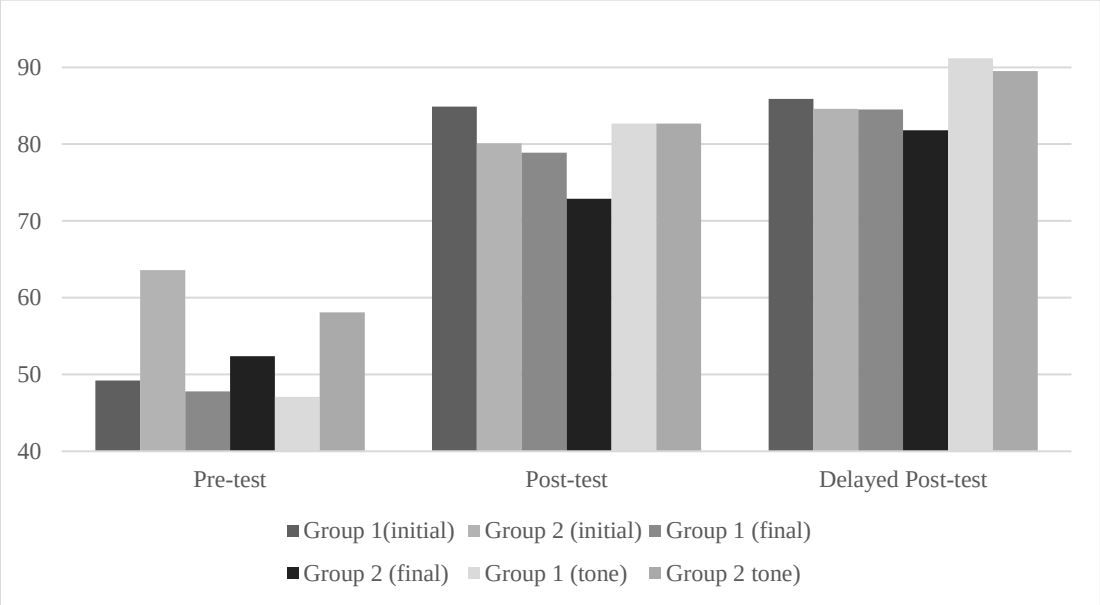

Figure 4 Production of Chinese syllables between the two groups using online applications

The descriptive data from their production of the Chinese syllables show a slightly different pattern than that of their perception. Both groups improved their pronunciation of initials, finals, and tones after the training, and retained their improvement or continued improving after the training. However, as indicated on their post-test, Group 1 showed greater improvement on their initials and tones than did Group 2. A mixed measures ANOVA was conducted with two within-group factors: time (three levels: pre-test, post-test, delayed post-test) and category (three levels: initial, final, and tone). There was no significant group main effect ($F(1, 15) = .066$, $p = .800$) or interaction effect among time, category, and group ($F(2.259, 12) = .758$, $p = .491$). However, there was a significant interaction effect of time and group: $F(1.304, 14) = 4.407$, $p = .040$. The results indicate that the Group 1 showed a significant improvement on the post-test than did Group 2.

In summary, both groups showed a similar pattern in their improvement in perceiving Chinese syllables, including initials, finals, and tones. However, Group 1 showed greater improvement in pronouncing Chinese syllables after the training, especially in pronouncing initials and tones. Group 2 did not improve as much as Group 1 but continued improving their pronunciation as shown on the delayed post-test.

5.4 Feedback from Participants on Their Experience with Online Applications

A survey was conducted to collect feedback from participants who received training using the online applications. Though the survey asked about participants' previous experience using online applications and computer technology in learning foreign languages, the main focus of the survey was on participants' experience with the online applications used in the present study, including their thoughts on the setup of the applications, their opinions about using these online applications, and their suggestions for improving their experience. See Appendix C for a complete list of the survey questions.

The analysis of their responses shows that about a quarter of the participants (23.5%) had never used any technology in learning a foreign language. Those who had used some technology, indicated that it was used occasionally. Only 30.8% of the participants indicated that they used some technology to learn foreign languages every day. Participants' experiences in using the online applications are summarized in Table 11. Overall, the participants had a pleasant experience with the online applications used in this study and were particularly happy with the instant feedback they received from both applications.

Table 11 Participants' feelings about using the online applications

Feelings about using online applications	Score*
Overall feeling	3.71
The set-up of online applications	3.65
The instant feedback provided	4.24

* The scores are based on a 5-point Linkert scale: from 1 being very unhappy to 5 being very happy

Participants' opinions on using online applications to learn Chinese syllables, including Pinyin, were also collected. They were asked to indicate to what extent they agreed with the statements listed below in Table 12 in which table their responses are also summarized.

Table 12 Participants' opinions on using the online applications

Statements	Score*
1. The online practice is very interesting.	3.41
2. The online practice is effective in helping me master pronunciation and Pinyin.	4.00
3. The online practice is hard for me.	3.18
4. The online practice helps me establish better connections between Chinese sounds and Pinyin.	4.18
5. The online practice familiarizes me with Chinese syllables.	4.06
6. I feel that, after listening to numerous Chinese syllables, I can pronounce them better.	3.82
7. I would recommend online practice to other learners of Chinese.	4.06
8. I hope to learn about other websites so that I can do more practice on my own.	4.18

*The scores are based on a 5-point Linkert scale: from 1 being strongly disagree to 5 being strongly agree

Overall, participants gave positive feedback on using online applications. They agreed that the online practice was interesting, easy, and effective in helping them master Chinese pronunciation and match sounds with Pinyin. They also perceived the online perception practice as helpful to their pronunciation. They would recommend using online applications to other learners of Chinese and would like to know about more online resources for learning pronunciation. The majority of the participants (88.2%) indicated that they either would or probably would use these applications in the future on their own.

Participants were also asked to indicate their preference between using online applications or using traditional class instruction to learn pronunciation. Their responses are summarized in Table 13.

Table 13 Participants' opinions on online applications vs. traditional instruction

Statements	Score*
1. I prefer online practice to in-class practice led by the instructor.	2.00
2. I prefer online practice to traditional workbook exercises after class.	2.71
3. I prefer using online practice in class with my teacher around to help.	3.47
4. I prefer using online practice on my own after class.	3.41

*The scores are based on a 5-point Linkert scale: from 1 being strongly disagree to 5 being strongly agree

The data show that participants were not enthusiastic about replacing in-class practice led by instructors with using online applications nor did they want to use online applications in place of traditional assignments outside of class. As for the context of using these online applications, whether in-class or outside of class, they showed a more or less neutral attitude.

An open-ended question was used to collect suggestions from participants about online applications. Some participants stated that they would like to hear about more online applications that work similarly. Some expressed a wish that the applications would provide easier access because the websites seemed a bit confusing at first. Some participants indicated that the pronunciation was hard to understand and would prefer for it to be articulated more clearly. One participant expressed a specific wish for a feature that would quiz users on the syllables they had had difficulty with.

6. Discussion

The goal of the present study was to investigate the effectiveness of online applications in helping L2 Chinese learners acquire the Chinese pronunciation system, including matching sounds with orthography. Specifically, the present study explored: a) if using online applications helped learners improve their perception of Chinese initials, finals, and tones; b) if online applications helped learners improve their production of Chinese syllables; c) if the context in which online applications were used had an impact on their effectiveness, and d) if online applications were well received by learners. It was not the goal of this study to explore whether online applications are significantly better than traditional instructor-led training and should then replace in-class instruction, but rather to explore their effectiveness and acceptance among learners so that they can be used as an extension of in-class training to meet the challenge of what has been identified as a lack of instruction time that can be designated for pronunciation training. In the following section, the results will be discussed based on the research questions the present study set out to answer.

6.1 Impact of Online Applications on Learners' Perception of Chinese Syllables

The first research question addressed the effect of online training in helping participants improve their perception of Chinese initials, finals, and tones. The results indicated that all the groups improved their perceptions of Chinese syllables significantly

after the training. There was no significant difference among the three groups' perceptions on either the post-test or the delayed post-test. However, the numerical results showed that the two groups that used online applications were better able to retain their improvement than the group that received traditional in-class instruction. The same pattern was found when their perception was tested using tokens that had been broken down into initials, finals, and tones. A similar pattern was also found in Xu et al. (2019), where the group that used an online application for tone perception training showed continued Tone 2 perception accuracy improvement whereas the group that received in-class traditional training stopping improving after the training session ended. It seems that the use of online applications has a more lasting effect than traditional in-class instruction. We would argue that this lasting effect is related to the fact that the two groups received more Chinese syllable input within the same amount of time (15 minutes) and also received more feedback on their practice from the online applications.

Some other common patterns were also identified. When focusing on the setting in which participants were required to provide the whole syllable – initials, finals, and tones – all of the groups had the most difficulty with the finals. Although the average perception accuracy of finals increased from 44.93% on the pretest to 71.21% on the post-test, and remained 71.55% on the delayed post-test, of the three categories, it still had the lowest accuracy rate. The group that used online applications in class achieved the highest accuracy on the delayed post-test (75.2%). Another related finding was that the setting of the task did have a significant impact on the accuracy of perceiving finals. On the delayed post-test, all groups did significantly better when the initials were provided than when they were not provided. Their accuracy was around 85% in the former setting and 71.6% in the latter. Participants' perception of initials and tones was more balanced in both settings. These findings are consistent with previous studies. Neri et al. (2006) found that L2 Dutch speakers' pronunciation of consonants improved without specific training, but their pronunciation of vowels did not improve at the same pace. Munro and Derwing (2008) discovered that L2 speakers of English with L1 Chinese or L1 Slavic had difficulty pronouncing English vowels even after being immersed in the target language for a long time. Bent et al. (2007) and Thomson (2011) both suggested that instruction in vowel pronunciation should be prioritized since vowels contribute more to speech intelligibility than consonants and are thus more beneficial in enabling learners to communicate intelligibly.

6.2 Impact of Online Applications on Learners' Production of Chinese Syllables

The second research question explored whether online applications, which are comprised of perception practices only, would also help improve participants' pronunciation of Chinese syllables. The results showed that all the three groups improved significantly with their pronunciation after the training, no matter what type of training they received. However, the numerical data from the post-test after training indicate that the group that received traditional instruction in class did better; they had higher scores in their pronunciation than the two groups that used online applications, especially the one that used online applications outside of class. However, by the time of the delayed post-test, the two groups that used online applications had caught up with and even exceeded

the performance of the traditional instruction group. The advantage of the traditional group on the post-test is not surprising: the participants did, after all, receive more production practice in class and more instant feedback from the instructor which clearly improved their pronunciation. However, the fact that this group failed to maintain its advantage on the delayed post-test seems to indicate that the online applications, used in class or outside of class, had a beneficial if delayed impact on participants' pronunciation.

Training effects from perception transferred to production have been found in previous studies as well (Bradlow et al.,1997; Thomson, 2011; Wang et al., 2003). An explanation for the observed transfer effect from perception to production could be that, with a large amount of input, learners develop a better perceptual system, which consequently enables them to better monitor their own production.

6.3 Impact of Context (in Class or outside of Class) on the Training Results

The third research question explored whether the context in which the online applications were used had an impact on the training effects. Results from both the perception task and the production task showed a very similar pattern of improvement over time in two of the groups which improved greatly on the post-test and increased their perception accuracy on the delayed post-test. However, when it came to pronouncing the Chinese syllables, the group that used the online applications in class showed significantly greater improvement on the post-test than the group that used the online applications outside of class. The significant difference was mainly seen in initials and tones. This finding is counterintuitive to some extent. The group that used online applications outside of class received traditional instruction in class, which included production practice and corrective feedback from the instructor. It was expected, therefore, that they would demonstrate better production than the group that used online applications in class, which included perceptual practice only. It could be a matter of skewed data due to the small number of participants in the group or it could be an indication of the strong effect transferred from perceptual training. Further research is needed to gain a better understanding of the exact cause. However, on the delayed post-test, the difference between the two groups disappeared: both groups showed improvement with the one that used applications outside of class showing even more. Previous meta-analysis studies on PI or CAPT compared the effect size in different training contexts. Lee et al. (2015) found that lab-based training had a greater effect than classroom-based training (0.95 vs. 0.79). Because, however, no studies, to the best knowledge of the authors, have explored the impact of the context of training involving an outside-of-classroom setting, no comparison of results can be made at this time.

6.4 Learners' Experience with and Perception of Online Applications

The last research question focused on participants' experience in using the online applications. The results from the survey showed overall positive feedback from the participants. They evaluated the perceptual practice with the online applications as moderately difficult. They felt that the practice helped them not only become more familiar with Chinese syllables and but also improved their production of same. Most of the

participants agreed that these online applications should be recommended to other L2 Chinese learners. They also showed interest in learning about other online applications designed to help them with pronunciation: about half of the participants set speaking skills as their primary goal in learning Chinese. They liked the instant feedback feature both applications provided, but expressed a wish for the applications to test them based on errors they made. It was also highly encouraging to discover that the majority of participants expressed a willingness to continue to use these applications in their future studies. At the same time, the majority of the participants indicated their preference for instructor-led in-class practice over online applications. The writers were not discouraged by this preference, since the goal of this study was not to recommend that in-class instruction be replaced but, rather, to find an effective way to extend in-class instruction.

7. Conclusion and Instructional Implications

Overall, the present study concludes that using online applications can help students improve their perception of Chinese syllables as effectively as traditional instruction and practice led by instructors in class. The groups that used online applications trended more toward gain retention and continued improvement on their perception of initials, finals, and tones after the training than the group that received traditional practice in class. All groups, regardless of the type of training they received, improved their production after training. The group that received traditional instructor-led instruction manifested more improvement right after the training than did the two online applications groups. However, this advantage was not retained. Whether the applications were used in class with an instructor's guidance or independently outside of class, no difference in the improvement of participants' perception was found. And, though using online applications in class resulted in better improvement in production right after the training, this advantage was not retained either after the training ended.

Based on the results from the tasks as well as an analysis of the participants' survey answers, the writers would suggest that online applications can be used outside of class as an effective supplementary aid for students in their quest to master the Chinese pronunciation system, including their perception of initials, finals, and tones. Although these online applications only provide perceptual practice, their high quality input can still help develop learners' perceptual system and thus their production. The present study only employed four 15-minute training sessions with the online applications over a two-week period. In a regular curriculum, it is recommended that the use of online applications be assigned for a longer period (30 minutes, for instance) and at a more frequent pace (three times a week, for instance). Based on participants' feedback, it is also suggested that students be provided with more options of similar online applications so that they can choose one that best fits their own needs or learning styles.

8. Limitation and Future Studies

Despite the authors' efforts to recruit as many participants as possible, the number participants involved was small and that was one of the limitations of the current study.

The exclusion of several participants due to incompletion of either a training section or a test further reduced the number. Furthermore, since this study used convenient samples, the distribution of participants in each group was not even. Caution must be exercised when interpreting the results, especially when the big variance within each group is taken into consideration. All of the above may have reduced the power to detect any significant difference among the groups. The findings may not be the same if a larger sample is used. Although this study included a comparison group, it was not a control group in the strictest sense since this group received traditional instruction. However, as part of regular instruction, it was impractical or, as Lee et al. (2015) pointed out, even "unethical to withhold treatment for the sake of experimental control" (p.363).

Another limitation that must be noted is the lack of strict control over the extra time the two online application groups may possibly have spent using those online applications on their own. At the beginning of the study, we encouraged both online application groups to complete, and only complete, the assigned tasks in order to make the time spent on Pinyin practice comparable among the groups. The group that were assigned online applications to use outside of class were asked to report the time they spent on those applications and they reported a time period similar to the required time 15 minutes each. This limitation, to some extent, resulted from the nature of the present study: a study conducted in a non-laboratory setting. In the future, maybe a lab-based training study can show the effects more precisely.

Two aspects can be further explored for future research. The first is the role of feedback embedded in online applications. As Rogerson-Revell (2021) points out, "accurate and timely feedback is essential" in CAPT to help learners acquire pronunciation. This was also echoed in participants' feedback when they indicated that the instant feedback was helpful. However, during the training, we noticed that some students showed frustration when, after several attempts, the feedback continued to indicate an error without pointing out which specific component was wrong. This begs the question: Would the training effect improve if more adequate and accurate feedback were provided? The second aspect to examine is the training on finals, the syllable segment which seems to be a stumbling block in both perception and production among L2 Chinese learners. The existing online applications mostly train learners' pronunciation holistically, without focusing on any specific segments. Many scholars have called for more collaboration "between pedagogic and technical experts" in designing CAPT tools (Rogerson-Revell, 2021). Thus, the kind of online applications that can be designed and employed to facilitate learners' acquisition of Chinese finals is certainly worthy of future consideration.

References

Arteaga, D. L. (2000). Articulatory phonetics in the first-year Spanish classroom. *The Modern Language Journal*, *84*(3), 339-354.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching*, *44*(3), 231-250.

Bent, T., Bradlow, A. R., & Smith, B. L. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In O. S. Bohn & M. J. Munro (Eds.), *Second language speech learning: The role of language experience in speech perception and production: A festschrift in honor of James E. Flege* (pp. 331-347). John Benjamins.

Beutner, M. (2001). Using computer assisted interactive feedback to enhance natural pronunciation of Chinese tones by non-native learners of Mandarin Chinese. [Unpublished doctoral dissertation]. Ohio University.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299-2310.

Chun, D., Jiang, Y., Meyr, J., & Yang, R. (2015). Acquisition of L2 Mandarin Chinese tones with learner-created tone visualizations. *Journal of Second Language Pronunciation, 1*(1), 86-114.

Collins, L., & Munoz, C. (2016). The foreign language classroom: current perspectives and future considerations. *The Modern Language Journal*, 100(S1), 133-147.

Everson, M. E. (2011). Best practices in teaching logographic and non-Roman writing systems to L2 learners. *Annual Review of Applied Linguistics, 31*, 249-274.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), S*peech perception and linguistic experience: Theoretical and methodological issues* (pp. 229-273). York Press.

Godfroid, A., Lin, C., & Ryu, C. (2017). Hearing and seeing tone through color: An efficacy study of web-based, multimodal Chinese tone perception training. *Language Learning*, *67*(4), 819-857.

Hu, M. (2009). Phonological awareness in Mandarin of Chinese and Americans (UMI 33665544) [Doctoral dissertation, Auburn University]. ProQuest Dissertations.

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal*, *27*(1), 1-12.

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In P. Li, L. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *The handbook of East Asian psycholinguistics: Volume 1, Chinese,* (pp. 209-217). Cambridge University Press.

Lambacher, S., Martens, W., Wakehi, K., Marasinghe, C., & Molhot, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26, 227-247.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*(3), 345-366.

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27, 184-202.

Li, Y., & Xu, H. (2018). Using online applications to help Chinese learners learn Pinyin [Conference session]. ACTFL 2018 Convention, New Orleans, LA, United States.

Lin, H., & Lin, C. (2010). Perceiving vowels and tones in Mandarin: The effect of literary phonetic systems on phonological awareness. In L. E. Clements & C.-M. L. Liu (Eds.), Proceedings of the 22nd North American Conference on Chinese

Linguistics (NACCL-22) & the 18th International Conference on Chinese Linguistics (IACL-18). Vol 1 (pp. 429-437). Harvard University.
Mahdi, H. S., & Khateeb, A. A. A. (2019). The effectiveness of computer-assisted pronunciation training: A meta-analysis. *Review of Education, 7*(3), 733-753.
Morin, R. (2007). A neglected aspect of the standards: Preparing foreign language Spanish teachers to teach pronunciation. *Foreign Language Annals, 40*(2), 342-360.
Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, 58, 479-502.
Neri, A., Cucciarini, C., & Strik, H. (2006). Selecting segmental errors in L2 Dutch for optimal pronunciation training. *International Review of Applied Linguistics in Language Training*, 44, 357-404.
Neri, A., Cucchiarini, C., Strik, H., & Boves, L. (2010). The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning, 15*(5), 441-467.
Olsberg, M., Xu, Y., & Green, J. (2007). Dependence of tone perception on syllable perception. *Proceedings of 8th Annual Conference of the International Speech Communication Association (Interspeech 2007)* (pp. 2649-2652). Curran Associates, Inc.
Pytlyk, C. (2011). Shared Orthography: Do shared written symbols influence the perception of L2 sounds? *The Modern Language Journal*, *95*(4), 541-557.
Rochet, B. L.(1995). Perception and production of second language speech sounds by adults. In W. Strange (Ed.), S*peech perception and linguistic experience: Theoretical and methodological issues* (pp. 379-410). York Press.
Rogerson-Revell, P. M. (2021). Computer-Assisted Pronunciation Training (CAPT): Current Issues and Future Directions. *The RELC Journal*, *52*(1), 189-205.
Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly, 46*(4), 842-854.
Sharma, B., Liu, C., & Yao, Y. (2015). Perceptual confusability of Mandarin sounds, tones, and syllables.Unpublished conference presentation. ICPhS, https://api.semanticscholar.org/CorpusID:31632891
Shei, C. (2014). *Understanding the Chinese language: A comprehensive linguistic introduction*. New York: Routledge.
Thomson, R., & Derwing, T. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3), 326-344.
Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, *28*(3), 744-765.
Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649-3658.
Wang, Y., Jongman, A., & Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone production before and after training. *Journal of the Acoustical Society of America*, 113, 1033-1043.

Xu, H., Li, Y., & Li, Y. (2019). Using online applications to improve tone perception among L2 learners of Chinese. *Journal of Technology and Chinese Language Teaching*, *10*(1), 26-56.

Zhang, H. (2018). Current trends in research of Chinese sound acquisition. In C. Ke (Ed.), *The Routledge handbook of Chinese second language acquisition* (pp. 217-233). Routledge.

Appendix A
Perception tasks used in the pre-test, post-test and delayed post-test

I. Please listen to the recording, and write down the initial and the tone for the syllables you hear. Each syllable will be read three times.

1. ___un	2. ___ai	3. ___ong	4. ___iu	5. ___ian	6. ___ i
7. ___ai	8.___iu	9. ___uo	10. ___u	11.___ao	12.___un
13.___eng	14.___ao	15.___ue	16. ___ian	17.___üe	18.___eng
19.___ong	20.___u	21.__uo	22. ___i		

Answers:
1. chun1 2. zhai1 3.nong2 4.xiu4 5.bian3 6.bi1 7.zai4
8. qiu2 9.zuo3 10. qu4 11.chao3 12.cun2 13. deng4 14. cao2
15. jue2 16.xian1 17. nüe4 18.reng1 19. zhong4 20.ju1 21.ruo4 22.di2

II. Please listen to the recording, and write down the final and the tone for the syllables you hear. Each syllable will be read three times.

1.b ___	2. q___	3. j___	4. r___	5. q___	6. x___
7. n___	8.b___	9. ch___	10. d___	11.z___	12.n___
13.z___	14.c___	15.r___	16. zh___	17.d___	18.ch___
19.x___	20.j___	21.c__	22. zh___		

Answers:
1. bi1 2.qu4 3.jue2 4.reng1 5.qiu2 6.xian1 7.nüe4
8.bian3 9.chun1 10.di2 11.zuo3 12.nong2 13.zai4 14.cun2
15.ruo4 16.zhai1 17.deng4 18.chao3 19.xiu4 20. ju1 21.cao2 22.zhong4

III. Please listen to the recording, and write the initial, the final and the tone for the syllables you hear. Each syllable will be read three times.

1. _____	2. _____	3. _____	4. _____	5. _____	6. _____
7. _____	8_____	9. _____	10. _____	11_____	12. _____
13_____	14. _____	15. _____	16. _____	17. _____	18. _____
19. _____	20. _____	21._____	22. _____		

Answers:
1.qian2 2. chi4 3.rong3 4.cuo1 5.nü3 6.jian4 7.zao1
8.cheng2 9.ci3 10.run4 11.jiu3 12.xu2 13.zong1 14.beng3
15.bai2 16.zhun3 17.diu1 18.xue3 19.zhao4 20.nuo2 21.dai3 22.que1

Appendix B
Production task

Please read and record the following syllables in order (1, 2, 3…). Save your recording by your full name as an .mp3 or .wmv file, and then submit the file to the Blackboard/Canvas site:

1.biǎn	2.zhāi	3.běng	4.diū	5.xú	6.zāo	7.jiǔ
8.cún	9.dèng	10.cáo	11.rēng	12.nóng	13.bī	14.jiàn
15.zuǒ	16.bái	17.zhào	18.cǐ	19.zhòng	20.nǚ	21.xiù
22.dǎi	23.rùn	24.quē	25. chǎo	26. nüè	27. cuō	28. qiú
29. zài	30. dí	31. ruò	32. rǒng	33. chì	34. nuó	35. jū
36. qián	37. chūn	38. xiān	39. xuě	40. jué	41. qù	42. chéng
43. zōng	44. zhǔn					

Appendix C
Survey used to collect participants' feedback on using online applications

Dear students, please share your experience on learning Pinyin with us. Your feedback is anonymous and will only be used to improve our curriculum. Thanks!

1.How long have you been using computer technology, i.e. computer, software, websites, and applications etc. in your foreign language studies? (circle one)

Never; less than 6 months; 6 months to a year; 1-2 years; 2-3 years; more than 3 years

2.How frequently do you use websites or applications to study foreign languages by yourself (not as part of a class)? (circle one)

Never; once a year; once a month; every 2-3 weeks; every week; 2-3 times a week; every day

3.How frequently do you use websites or applications to study foreign languages as part of a class?

Never; once a year; once a month; every 2-3 weeks; every week; 2-3 times a week; every day

4.You have been required to use some websites (www.pinpractice.com; https://sla.talkbank.org/pinyin/) to practice Pinyin recently. Do you think you will continue to use these websites on your own in the future when you are not required to do so?

Yes; Maybe; Probably not; No

Why?_____.

5.What Chinese skills are you most interested in learning?

Listening; Speaking (including Pinyin); Reading; Writing (including characters)

6. **According to your experience, using the websites** (www.pinpractice.com; https://sla.talkbank.org/pinyin/) **to practice Pinyin, makes you feel** (circle all that apply):

Very happy/ somewhat happy/ neither happy nor unhappy/ unhappy /very unhappy

- **The set-up of the online exercises** (the format of the practice, the way to input your answer, etc.) **makes you feel** (circle all that apply):

Very happy/ somewhat happy/ neither happy nor unhappy/ unhappy /very unhappy

- **The instant feedback you get after you input the answer, makes you feel**:

Very happy/ somewhat happy/ neither happy nor unhappy/ unhappy /very unhappy

- **The holistic experience of using a website to practice Pinyin, makes you feel**:

Very happy/ somewhat happy/ neither happy nor unhappy/ unhappy /very unhappy

7. **To what extent do you agree with the following statements?** (circle one)

   - **The online Pinyin practice is very interesting**.

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

   - **The online Pinyin practice is efficient in helping me master Pinyin.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

   - **The online Pinyin practice is very hard for me.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

   - **The online Pinyin practice helps me establish better connections between Pinyin and Chinese sounds.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

   - **The online Pinyin practice familiarizes me with Chinese syllables.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

   - **I feel that, after listening to numerous Chinese syllables, I can pronounce Pinyin better**.

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

   - **I like the feedback I get in the Pinyin online practice.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

- **I don’t think feedback is necessary in the online Pinyin practice.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

- **I prefer online Pinyin practice to in-class practice led by the instructor.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

- **I prefer online Pinyin practice to traditional workbook exercises after class.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

- **I prefer using online Pinyin practice in class with a teacher around to help.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

- **I prefer using online Pinyin practice on my own after class.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

- **I would recommend online Pinyin practice to other learners of Chinese.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

- **I hope to learn about other websites so that I can do more Pinyin practice on my own.**

Strongly disagree (1), Disagree (2); Neither agree nor disagree (3); Agree (4); Strongly agree (5)

8. What change would you like to see to improve your experience with online Pinyin practice?