【附件三】教育部教學實踐研究計畫成果報告(系統端上傳 PDF 檔)

教育部教學實踐研究計畫成果報告(封面)

Project Report for MOE Teaching Practice Research Program (Cover Page)

計畫編號/Project Number：PHA1090439

學門專案分類/Division：人文藝術及設計

執行期間/Funding Period：2020/09/01-2021/8/31

透過運用情感反應語言教學（ERLE）來增進學生的英語交流和溝通能力

英語發音與聽力練習/英語會話與溝通

計畫主持人(Principal Investigator)：Beate Luo*

共同主持人(Co-Principal Investigator)：J. Sloan**, J. Carson-Berndsen**

執行機構及系所(Institution/Department/Program)：

*: Department of Foreign Languages/Feng Chia University

**: School of Computer Science/University College Dublin

成果報告公開日期：

■立即公開 □延後公開(統一於 2023 年 9 月 30 日公開)

繳交報告日期(Report Submission Date)：

2021 年 9 月 30 日

Enhancing students' English communication skills through Emotional Response Language Education (ERLE)

一. 報告內文(Content)(至少 3 頁)

1. 研究動機與目的(Research Motive and Purpose)

The motive of this study was to test if the learning platform ERLE can improve students' English grammar and pronunciation through a conversational setting while receiving native speaker feedback. Results of this study are also used by University College Dublin, where the learning platform is being further developed, to receive further funding for this platform so that it can be offered open and for free for everyone in the future which will benefit all of our students who want to use the platform after their freshman year.

The purpose of this research project was to improve students' ability to express themselves in English. Both aspects, pronunciation and grammar, are of great importance. As I have observed over the many years I have been teaching in Taiwan, many students who graduate from an English or Foreign Language Department still use incorrect grammar and speak with a strong accent which sometimes makes it difficult to understand what they want to communicate. Often, their teachers have the same pronunciation problems, which makes it difficult for their students to get aware of their own problems or to get valuable feedback on their pronunciation. As for grammar problems, as teachers we again and again find the same problems in students' spoken or written productions of English. Changing teachers from one course to the next does not provide consistent feedback as different teachers set different targets for their students. Thus, a learning platform where students can practice speaking English outside the classroom with native speaker feedback on grammar as well as pronunciation facilitates life-long learning which is necessary in the case of changing

fossilized errors. Once the platform is running and can prove to have a positive effect on students' English language skills, students will be able to use it as long as they want – after finishing their first year or even after graduation. Furthermore, our students often expose language anxiety when they speak English. Thus, we want students to experience the possibility to practice the specific problems they have in a non-threatening environment while receiving native speaker feedback and hope that, at the same time, they will get more confident when they need to use English in a – for them – more threatening environment.

2. 文獻探討(Literature Review)

Many Taiwanese children start learning English already at kindergarten. Thus, they would still be able to develop a native speaker-like pronunciation. However, due to the emphasis on computer-based testing in the educational system, pronunciation training is utterly neglected and students 'inherit' their local teachers' pronunciation problems (Luo, 2016). These problems become fossilized over time, meaning they are highly resistant to change (Selinker, 1972). Using peer reviewed pronunciation reading (PPR), Luo (2014) showed that the more students became aware of the gap between their peers' pronunciation and the target and gave feedback to their peers, the more they were able to improve their own pronunciation as they also noticed the gap between their own pronunciation and the target. Furthermore, they improved their pronunciation to a higher degree than students who only received in-class pronunciation training (Luo, 2016).

Pronunciation problems as well as grammar problems are mostly due to an influence of the L1 and are thus typical for language students of a specific region. Typical pronunciation problems of Taiwanese students have been described by Luo (2016) and

grammar problems by Chen (2002) and Chang and Tsai (2007). The most common problems for pronunciation were incorrect phonemes, omission of endings, adding an extra syllable or deleting a consonant to simplify consonant clusters, linking words in thought groups, word and sentence stress, as well as intonation. The most common grammar problem were word usage, tense, article usage, use of prepositions, verb forms, use of singular/plural, relative clauses, awkward sentences, and redundancies.

However, correction rates for student errors by language teachers is generally low (Chaudron, 1988; Carroll & Swain, 1993) as many teachers try to avoid explicit, negative verbal feed-back in the classroom and only use non-verbal cues extensively (Wang & Loewen, 2016). The reasons for teachers not to correct their students range from not wanting to disrupt the flow of conversation to believing that correction may hinder acquisition (Truscott, 1996). On the other hand, there is little information on the effect of multi-modal feedback through non-verbal cues in the literature (Lyster, Saito & Sato, 2013).

The learning platform ERLE (Emotional Response Learning Environment) provides students with an engaging and immersive virtual interaction (Sloan & Carson-Berndsen, 2018). Different from deferred feedback as offered by classmates in PPR (Luo, 2014 and Luo, 2016), ERLE provides immediate feedback by a native speaker through changes in facial expressions and gaze of the human-like avatar. This form of feedback causes language students to reflect on their production and leads to less complex further utterances (Sloan & Carson-Berndsen, 2017). However, as Arnold (2011) pointed out, just as negative affect can hinder or even prevent learning from taking place altogether, positive affect can support learning invaluably. Therefore, the human-like avatar in ERLE was developed to not only show confusion when errors

occur but also to give positive feedback by nodding and smiling when input was correct as well as reacting to the content of the input by showing disgust, fear, sadness, interest, happiness, and surprise. The user interface is a crucial factor when giving language students explicit corrective feedback on their errors as correction can lead to embarrassment, frustration, and anxiety (Sloan & Carson-Berndsen, 2018). Language anxiety is the apprehension students experience when they need to use a foreign language without being fully proficient (Gardner & MacIntyre, 1003) and is a learned emotional response to repeated negative experiences (MacIntyre & Gardner, 1989). It has been widely reported in language students from East Asia (Woodrow, 2006). However, interaction between the student and an avatar can decrease language anxiety as it reduces fear of making mistakes or not knowing what to say next on part of the student (Sloan & Carson-Berndsen, 2018).

The interaction with the avatar on ERLE was designed with the aim of informing the student, with the least amount of written and spoken instruction, how to use the interface. This is in line with the principle of minimizing the users' cognitive load through human-centered design (Oviatt, 2006). Consideration for reducing this load is of even higher concern when the end users are intermediate-level ESL students who are already preparing for a demanding task – speaking in English. Where possible, easy to understand signifiers (Norman, 1988), e.g. microphone and square stop symbols, were included to be understandable by users regardless of first language. As non-linguistic related errors (i.e. by pressing the wrong button) might cause a negative experience, it was necessary to test, if the created interface on ERLE was easy-to-understand and highly usable. This was done using heuristic evaluation (Sloan, Luo, & Carson-Berndsen, 2019) whereby users explore and evaluate the interface on their

own terms (Tan, Liu, & Bishu, 2009). Heuristic testing is relatively quick to implement and can be used early in the design process. Nielsen & Molich (1990) tested four different interfaces and were able to identify 55 - 90% of usability problems through heuristic testing by just five evaluators.

The 'in-the-wild', heuristic user evaluation of the ERLE platform with five English as a Foreign Language students from Feng Chia University in Taiwan and one native English speaker in Ireland was performed over three months. The feedback garnered led to the introduction of a tutorial prior to the initial class, a redesign of the buttons and presentation of the Automatic Speech Recognition (ASR) output, and an animated response to loud input which causes difficulty for the ASR system (Sloan et al., 2019). The improved system was then implemented to a first-year English Speaking and Listening course at the same university in a larger, longitudinal study. Results of this study have let to further improvements of the platform. With the new version, students can immediately reflect not only on their grammar but also on their pronunciation errors and try again to correctly pronounce the incorrect utterance. This kind of explicit feedback has a significant positive effect on language learning (Li, 2010; Lyster, et al., 2013; Mackey & Goo, 2007).

3. 研究問題(Research Question)

Due to the financial problems in Ireland caused by the COVID 19 health crisis, the Spectropin feature could not be further developed by UCD and was, thus, not used in this study. Instead, the improvements in students' pronunciation were compared to results from another group, where students had used Peer reviewed Pronunciation Reading (PPR). Therefore, the research questions had to be slightly modified.

As mentioned in the literature review, improvement in the frequency of errors relies on

the students noticing the gap between their interlanguage and the target. The error signal used in the interaction on ERLE provides students with the opportunity to do this. The reduction of sentence complexity and length after viewing this signal was evident in earlier testing of the site (Sloan, pers. communication). This behavior pointed to a more cautious approach to subsequent sentences. Therefore, it would be predicted that learners who use ERLE in this experiment would also become more cautious in their production. This caution could be measured through frequency of hesitations and repetitions. Thus, a further research question was added as follows:

1. Does the use of the ERLE platform lead to an increase in the frequency of hesitations and repetitions in students' speech?
2. Does the use of ERLE help students to improve their pronunciation?
3. Do improvements differ from those of students who use Peer Reviewed Pronunciation Reading (PPR)?
4. Does the use of ERLE help students to reduce their grammatical and lexical errors in their oral language production?
5. Are improvements in the use of grammar and lexis in oral language production also reflected in students' written language production after an extended period?

The underlying hypotheses where that the use of ERLE can reduce the frequency of errors in students' oral production in terms of pronunciation (H_1) and grammar/lexis (H_2), that improvements in pronunciation are comparable to improvements achieved through PPR (H_3) and that improvements in grammar/lexis will still be noticeable after an extended period (H_4)

4. 研究設計與方法(Research Methodology)

(1) Participants

A total of 68 participants were recruited from the Department of foreign Languages and Literature of Feng Chia University. In the first week of the fall semester 2020, a demonstration video of ERLE and a description of the experiment were presented to students taking the 'English pronunciation and listening' courses offered in the first semester of their freshman year. Students' English ability ranged between intermediate to high intermediate (B1 or B2 on the Common European Framework of Reference for Languages). The aims and objectives of the experiment were explained to the students Volunteers signed up to the ERLE platform and took a pre-test. After one week, 56 students had successfully completed the pre-test. The remaining 12 did not attempt and were excluded from the experiment. The 56 students who successfully completed the pre-test were randomly split into an experimental group and a control group. Two students in the control group did not take the posttest after the final round of ERLE sessions and were also excluded from the experiment. The experimental group thus contained 28 students, and the control group 26. However, only 22 students in the experimental group and 25 students in the control group provided recordings for both reading assignment. The others either recorded only one of the readings or one of the recordings was of too poor quality to be evaluated.

Results from the experimental group were also compared to results of students who had performed PPR to improve their pronunciation in the previous year. In that group, 22 students had volunteered to allow their data to be used in this study.

(2) Experimental design (pronunciation)

An independent measures design was used for this experiment. The experimental group

used ERLE once a week, for 30 minutes each time, over an 8-week period. The control group did not. In addition to the pre and post tests taken before and after the 8-week experimental period, the experimental group completed a questionnaire after finishing their 8th class (see chapter 4.4). Pronunciation problems were rated by the researcher and a trained assistant according to the rubric in Table 1. In case of differences between the two raters, results were discussed to come to a final conclusion.

Table 1: Grading rubric for pronunciation problems (Luo, 2014)

severity / level	0 no errors, completely comprehensible	0.5-1 minor errors that do not affect intelligibility	1.5-2 several errors that might affect intelligibility	2.5-3 severe errors that affect intelligibility
phoneme	no errors	minor error in one or two phonemes	minor error in more than two phonemes or more severe errors in at least one phoneme	severe errors in more than one phoneme
cluster	no errors	occasional problems with clusters	more often problems with clusters	majority of clusters is mispronounced
word	correct word stress and no omission of syllables	occasional incorrect word stress or omission of syllables	more frequent incorrect word stress and/or omission of syllables	word stress frequently incorrect and syllables often omitted
sentence	very good linking, sentence stress and intonation	good linking, sentence stress and intonation	some variation in intonation and sentence stress, linking sometimes a problem	words are seldom to never linked, little to no variation in sentence stress and intonation

For comparing differences in students' pronunciation with results from students who had been using PPR, another pre- and posttests was used in the first semester in addition to the pre-and posttest in ERLE. Students were asked to record the same texts

as in the PPR experiment at the beginning and the end of the semester, respectively. Over a period of eight weeks, students in the PPR group had shadow read a new poem each week to practice correct pronunciation, recorded the poem, and uploaded it to a learning management system (LMS). They then listened to the recordings of three peers and provided written feedback on pronunciation problems in the LMS. For a detailed description, see Luo (2014 & 2016).

(3) Experimental design (grammar)

In this part, students' oral and written production of English was tested. A pre-/post-/delayed posttest design was applied whereby the pretest consisted of the transcriptions of students oral pretest which had been included in the ERLE web application and was completed a week before conversation sessions in ERLE started. The posttest also consisted of the transcriptions of a similar test students took the week following the end of conversation classes in ERLE. The transcripts were provided by the co-researchers from UCD. In these tests, students were given 30 seconds each to answer three different questions related to familiar topics. Three sets of questions were created, with each participant getting a random choice when they began the pretest. The posttest was then randomly selected from the remaining two.

The pre- and posttest question sets are shown below:

• Let's talk about your hometown

Q1. What kind of place is it?

Q2. What kind of jobs do people in your town do?

Q3. Is it a good place to live?

• Let's talk about your university

Q1. What kind of place is it?

Q2. What kind of students study at your university?

Q3. Is it a good place to study?

• Let's talk about vacations

Q1. Where do you usually go?

Q2. What kind of people do you meet there?

Q3. What is your favorite thing about that place?

The delayed posttest was the writing proficiency test students took at the end of the second semester. In this test, students were given 50 minutes to write an essay on the topic: "An event that shaped my life". All students from the experimental and control group who participated in the second part of this research investigating the influence of ERLE on students' grammar improvements received additional grammar training in their respective writing classes over the two-semester period.

Evaluation of grammar problems were rated by the researcher and a trained assistant, independently from each other. In case of differences between the two raters, results were discussed to come to a final conclusion.

(4) Experimental Condition

Students in the experimental group took one 30-minute class per week over an 8 week period. Time difference between Ireland and Taiwan dictated that took classes in the evening, which was the late morning/early afternoon in Ireland. Classes were generally open between 7pm and 10pm (Taiwan Standard Time) on weekdays, and the students could choose when to log in and take classes. Upon entering class, students were asked how they were feeling and what they would like to talk about in the class.

The students then spoke to the teacher by recording their voices, checking the ASR transcription and submitting the sentence. An example of this process, where the

student attempted to say, “I think stay in home is very bored” is shown in Figure 1. The student in this case tried three times to record this sentence. The first two times the word “sing” was shown instead of “think”. The third time the transcription displayed "sync" (a). The participant decided not to try and record again, but tapped the word (b). This opened an interface where it was possible to type the target word (c), which then replaced the non-target word in the transcription (d). The participant then submitted this sentence. This example is common, as many of our students begin sentences with “I think...” and Chinese L1 ESL students frequently have difficulty with fricatives and final plosives. “I sing” was often followed by "I sync" or "I thing" as students appeared to notice a deficiency in their pronunciation and attempted to fix it. The typing interface allows students to submit their desired sentence, even when they can’t achieve the target through speaking. The pattern of behavior showed that most students would re-record once or twice before typing, which could still have a positive effect on pronunciation - even if only in the consciousness-raising effect of realizing that certain words are frequently misunderstood by the system.

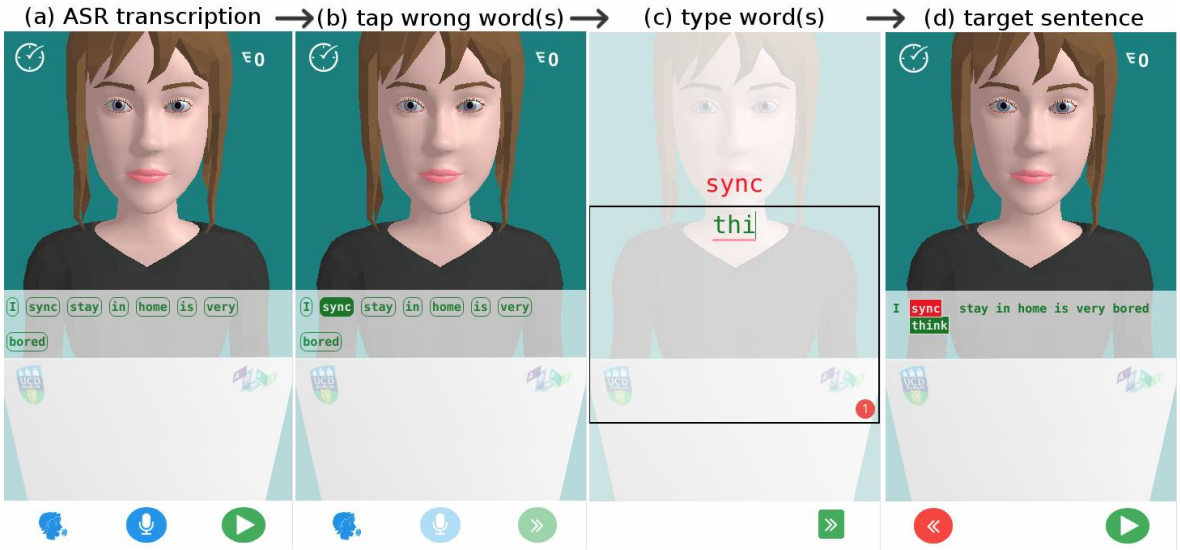


Figure 1: Sequence of ASR Interaction

Once the sentence was submitted, the avatar, Saoirse (pronounced Seersha, an Irish name), thought through the sentences and then responded in one of two different ways depending on whether the sentence was well or ill-formed. For a well-formed utterance, her facial expressions changed in accordance to the content of the sentence as judged by the teacher (one of the co-researchers from UCD), with a spoken prompt to follow. This process was aided by automation to allow several students to take classes at the same time.

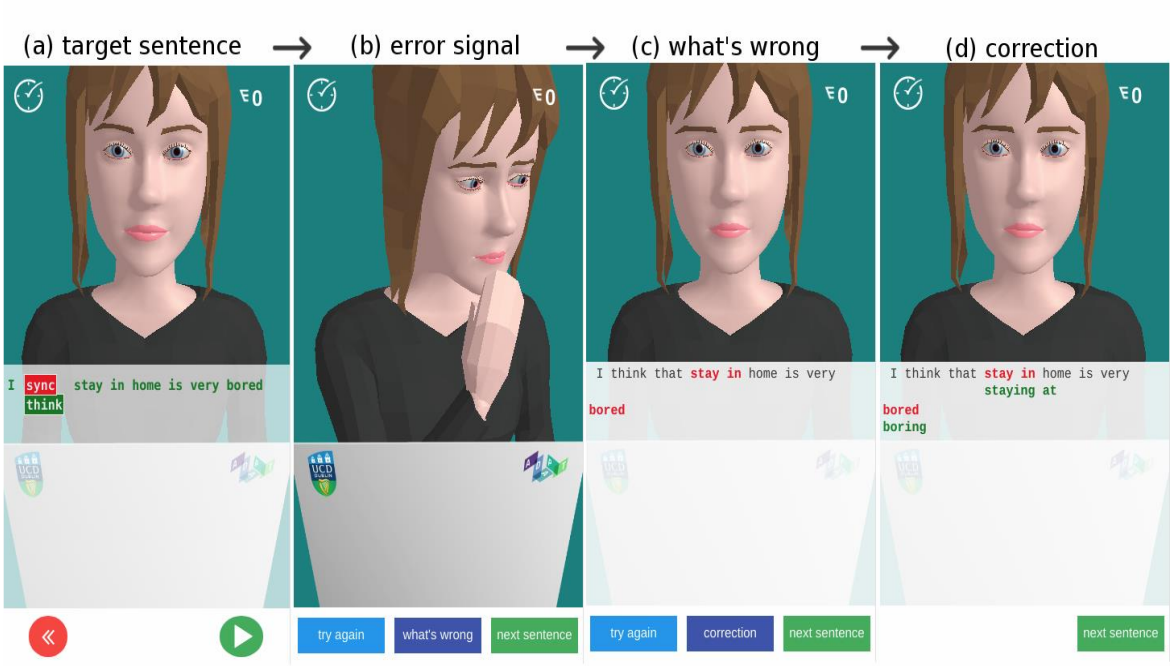


Figure 2: Sequence of Error Interaction

For an ill-formed utterance as in Figure 2a, after thinking through the sentence, Saoirse responded with the paralinguistic error signal used consistently in this research - gaze aversion and a frown Figure 2b. This began the error interaction. If the student selected the what's wrong button, the location of errors were shown in bold and red. In this case, "stay in" and "bored" were highlighted as shown in Figure 2c. By tapping show correction, one possible correction, as input by the teacher, was displayed. For this example, "staying at" and "boring" appeared below the locations of errors in Figure 2d.

The student could also have chosen to try again or simply move on to the next sentence at (b) and (c), without viewing the location of error(s) or correction(s). As noted by the co-researchers of UCD throughout this research, giving the students choice and autonomy over what happens after they receive the error signal is a crucial element of this error interaction.

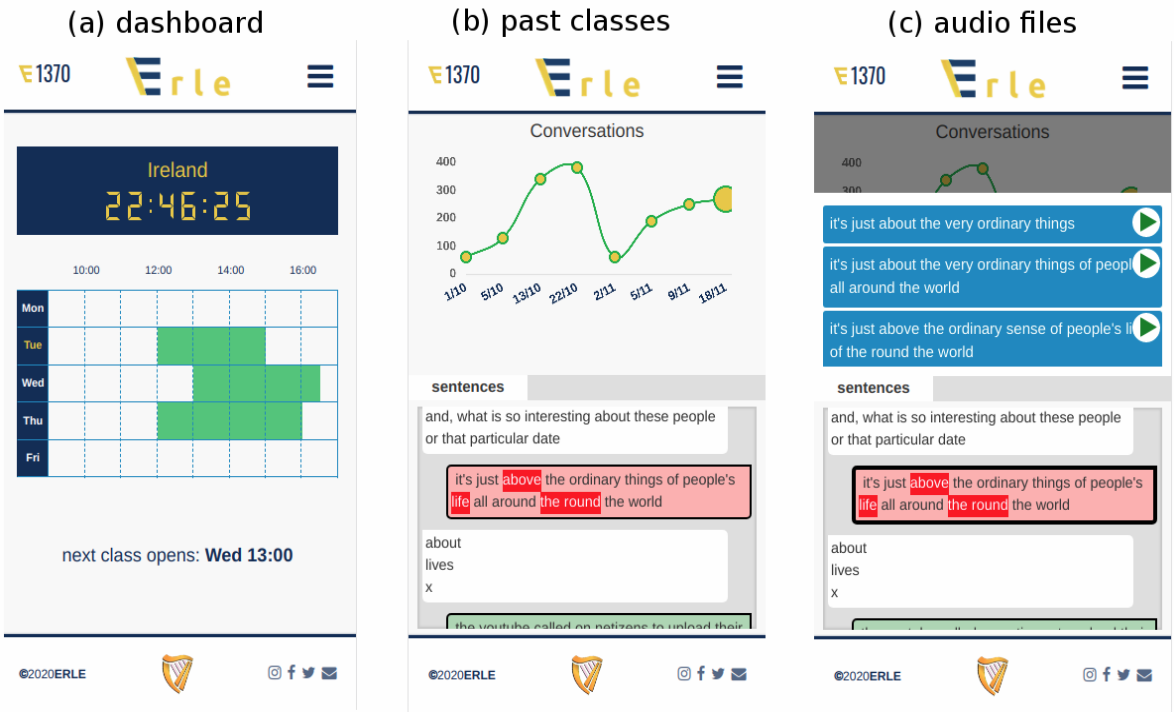


Figure 3: Student dashboard in ERLE

After 30 minutes, students were prompted to finish the class. Upon exiting class, an interface showing the full conversation was displayed (see Figure 3b). Points were awarded for correct sentences (10 for short, 20 for medium and 30 for long sentences) and were plotted on a graph at the top. All previous classes could be accessed by tapping on the points on the graph. By tapping individual sentences, students could listen to all the recordings they made and see the resulting, unaltered transcriptions (Figure 3c). This functionality allowed them to review their classes, sentences and errors to aid in their improvement. Students could also review on the dashboard when classes would be open (Figure 3a) whereby the schedule shows Ireland time.

(5) Questionnaire

After completing the eight weeks of classes with Saoirse, students in the experimental group answered a detailed questionnaire on their experiences with the ERLE platform over the research period. Usability was measured through a number of general factors, including performance expectancy, hedonic motivation and behavioral intention. The questionnaire consisted of 51 questions, 3 of them elicited explicit answers while the rest were given on a 5-point Likert scale. Students were also given the option to add additional remarks. The questionnaire and the respective responses from students are shown in the Appendix (A).

5. 教學暨研究成果(Teaching and research outcomes)

(1) 教學過程與成果 Teaching process and results

A. Frequency of repetitions and hesitations

Hesitations and repetitions are an indicator for the need to think before talking. Thus, an increase in both features had been expected, as students might get more cautious when speaking. However, there were no differences in the percentage of hesitations neither in the pretest and the posttest for each group nor between groups, due to the large standard deviations. Furthermore, even though these differences were not significant, different from expectations, reduction in repetitions was more than 7 times as high in the experimental group compared to the control group.

Similarly, due to the large standard deviations, no effect was found in the reduction in frequency of hesitations given in percent for the experimental group over the control group ($p=0.076$). However, the mean frequency of hesitations in the experimental group decreased significantly from pre- to posttest and differences between groups became significant at $p=0.015$ in the posttest (Table 2). Thus, instead of becoming more

cautious in their speech, students in the experimental group reduced slightly repetitions and hesitations significantly, thereby indicating an increase in their fluency of speech. Several of students' comments also indicate that students realized that their speaking ability improved. For example, one student stated s/he felt less nervous when speaking. Others indicated that it was helpful to have additional chances to use English outside classes (see Appendix (A)).

Table 2: Frequency of hesitations and repetitions given in percent to total word count

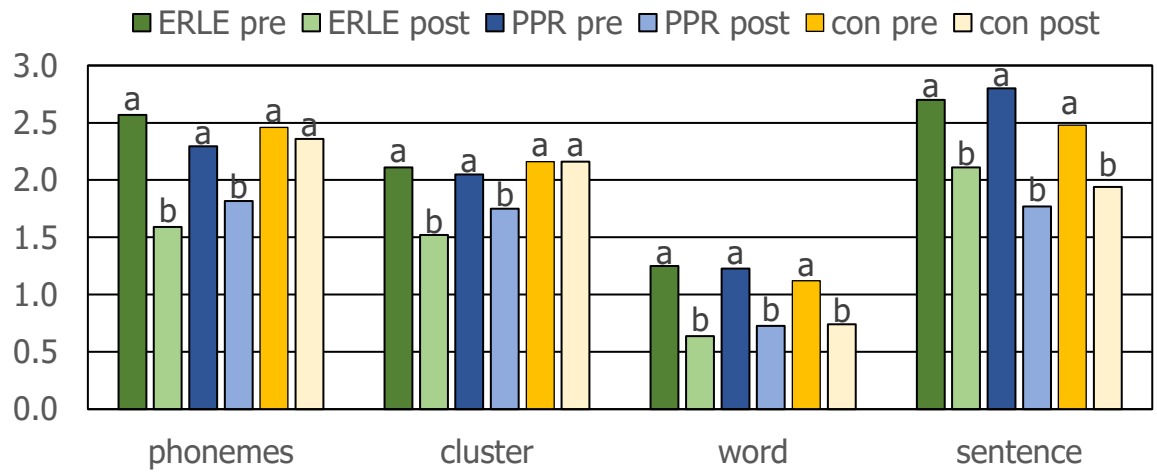
item			pretest	posttest	difference	significance
frequency of repetitions	experimental	mean	4.74	3.88	-0.96	0.145
		SD	4.41	3.14	4.74	
	control	mean	4.56	4.43	-0.13	0.906
		SD	4.00	4.24	5.46	
	significance	(p)	0.810	0.291	0.275	
frequency of hesitations	experimental	mean	5.57	4.30	-1.63	0.014
		SD	3.64	2.42	3.68	
	control	mean	6.18	6.07	-0.11	0.885
		SD	3.05	3.36	3.95	
	significance	(p)	0.786	0.015	0.076	

B. Pronunciation

Figure 3 shows the intensity of pronunciation problems in the pre- and in the posttest. The respective data are given in Appendix (B). There were no differences between groups in the pretest in any of the different problems tested. However, students in the ERLE and in the PPR group tested significantly better on the phoneme and cluster level, while all students improved on the word and sentence levels.

On the other hand, when observing the differences between pre- and posttests, (Figure 4), differences between groups become more differentiated, as slightly higher results in the pretest and slightly lower results in the posttest, differences between groups become more obvious. To test for significances, T-Tests were performed between the

different groups and revealed significantly higher improvements for ERLE (M=-0.98, SD=0.24) compared to PPR (M=-0.48, SD=29, p=0.000), and for both treatments compared to the control (p=0.000 for ERLE/control and p=0.000 for PPR/control). These differences are marked through different letters on the columns for each group. On the cluster level, differences between groups are comparable to the phoneme level. Reduction in the ERLE group were highest (M=-0.59, SD=0.40) and significantly higher compared to PPR (M=-0.30, SD=0.33, p=0.005). In the control group, there were no improvements at all (M=0.00, SD=0.41) and differences were significant between ERLE and control (p=0.005) as well as PPR and control (p=0.008).



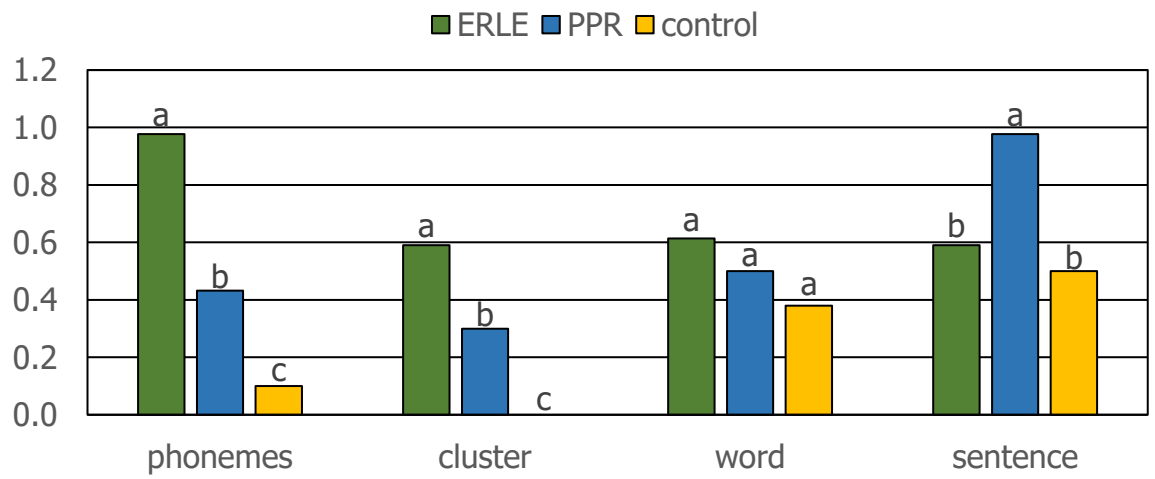
Significant differences between groups at p<0.5 are indicated by different letters.

Figure 4: Intensity of pronunciation problems in the pre- and posttests

Differences in improvements on the word level were not significant between groups with p=0.281 (ERLE/PPR), p=0.238 (PPR/control), and p=0.097 (ERLE/control). On the other hand, differences were significantly higher in the PPR group (M=-1.02, SD=0.42) compared to ERLE (M=-0.59, SD=0.43, p=0.001) and to the control (M=-0.50, SD=0.50, p=0.000), while differences between ERLE and the control were not significant (p=0.254). Thus, ERLE was more effective in improving sounds on the phoneme and cluster level compared to PPR while PPR was more effective on the

sentence level. These results make sense, as students received immediate feedback in ERLE so that they could rerecord incorrectly pronounced words, thus making them aware of their pronunciation problems (compare students remarks in Appendix (A)). On the other hand, the ASR transcribing students' utterances cannot distinguish between good and poor linking, intonation or incorrect word stress. In this case, feedback in PPR, although delayed, was more helpful. Thus, improvements in the ERLE group were more likely due to in-class instruction.

Neither PPR nor ERLE were more effective on reducing problems on the word level. Different from earlier studies (Luo, 2014, Luo 2016) problems on the word level were mostly due to word stress while few students omitted and even less students added extra syllables. However, word stress in English is difficult to predict. Thus, improvements in all three groups seem to stem mostly from class instruction alone.



Significant differences between groups at $p<0.5$ are indicated by different letters.

Figure 5: Reductions in the intensity of pronunciation problems

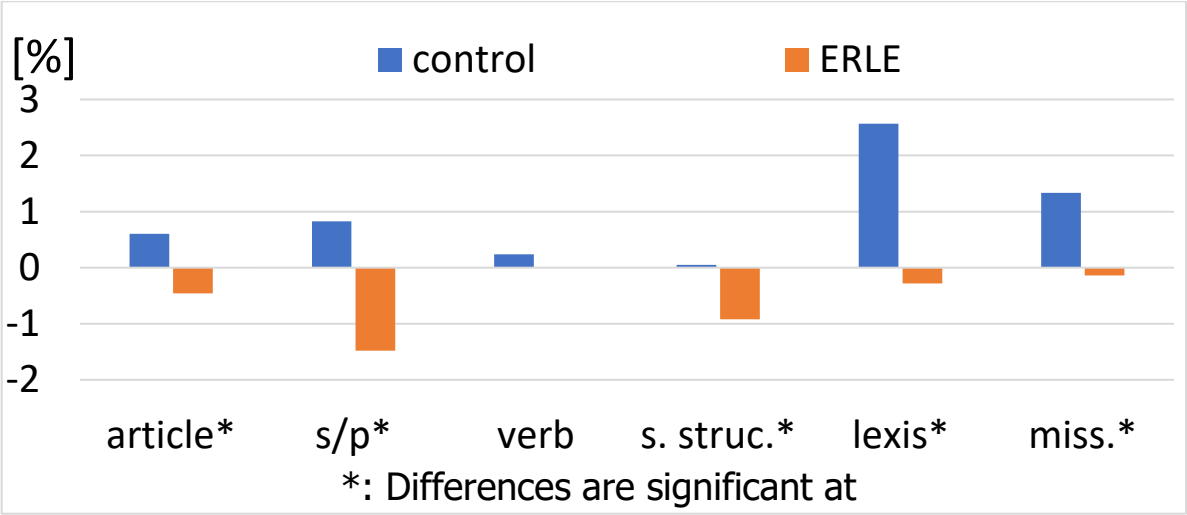
Reviewing the data from Luo (2016), it became clear that omission of syllables had been a much more severe problem compared to the last few years. Improvements in the experimental group on the word level at that time depended more on a reduction of omission of syllables and to a far lesser extend on the improvements on word stress.

Improvements in one of the control groups on word level, however, were due to the special emphasis the teacher of that group had given to word level problems. Predicting word stress is difficult for learners of English as there are too many exceptions (e.g., the **no**tice – to **no**tice) to the many rules (e.g., the **re**cord – to re**cord**). However, omission of syllables has become a much less severe error problem in the last few years. Consequently, we can accept our first hypothesis (H_1) that ERLE reduces the frequency of pronunciation errors in students' oral production. However, it is superior compared to PPR in terms of phoneme and cluster problems while PPR is superior on the sentence level (H_3).

C. Grammar and lexical errors

(a) Oral language production

Figure 6 shows the differences in the intensity of grammar and lexical errors in the oral pre- and posttest. The respective data are given in Appendix (C). Compared to the pretest, students in the experimental group improved significantly on the use of singular/plural and sentence structure. The other improvements were too small to be significant. However, as all students had received grammar instruction in their respective writing classes, the significant increase in grammatical and lexical errors in the control group on articles, lexis, and missing words was unexpected. They led to significant differences in the magnitude of differences between the ERLE group and the control group and even the slight improvements in article, lexis, and missing word errors in the experimental group became significant compared to the increase of the respective errors in the control group (Figure 6).



s/p=singular/plural, s. struc.= sentence structure, miss.=missing words

Figure 6: Differences in the intensity of pronunciation problems

These results let us accept the second hypothesis (H_2) that ERLE reduces the frequency of grammatical and lexical errors in students' oral production.

(b) Written language production

Students took the writing proficiency test six months after finishing the oral posttest in ERLE. All students had received grammar training as part of their Writing 1 and Writing 2 courses. Nonetheless, differences observed in the posttest in oral language production (Figure 4, Table C in Appendix) were still visible in their written language production except for sentence structure where students from the experimental group again produced a comparable percentage of errors. On verb errors, on the other hand, students in the experimental group still did better than their peers in the control group, however, the percentage of errors was much higher compared to their oral language production (Figure 7, Table 3). It has to be kept in mind, though, that in the oral tests, students could answer the questions without the need to change tense. On the other hand, in the written delayed posttest they were asked to write about an event that had happened in the past and changes in tense are error prone for our students as it is not

comparable with Chinese. Still, students of the experimental group performed better compared to their peers from the control group.

Differences in grammar problems in the delayed posttest

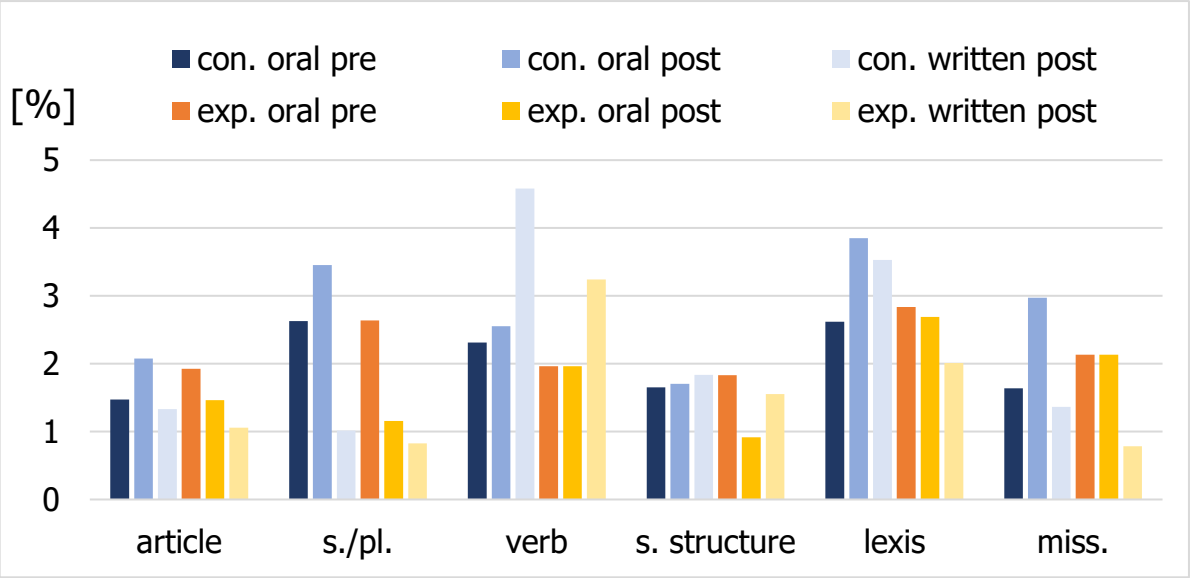
		ERLE	control	significance (p)
articles	mean	1.06	1.33	0.162
	SD	0.70	1.01	
singular/plural	mean	0.83	1.02	0.239
	SD	0.73	0.91	
verb	mean	3.24	4.58	0.040
	SD	1.54	2.90	
sentence structure	mean	1.55	1.84	0.226
	SD	0.93	1.37	
lexis	mean	2.01	3.53	0.005
	SD	1.56	1.96	
missing words	mean	0.79	1.37	0.028
	SD	0.75	1.07	
total errors	mean	11.34	16.28	0.007
	SD	4.26	7.30	

Furthermore, in written language production, students are used to write longer sentences, as this had been required for their Chinese written language production since elementary school. Thus, they also tend to write longer sentences in English and the longer the sentences, the higher the probability to produce errors. It seems that the reduction in sentence length of almost 10% after an error signal in the previous sentence (Sloan, personal communication) might have influenced the reduction in sentence structure in the oral posttest. However, these signals were missing in the written delayed posttest.

The reduction in errors on singular/plural observed in the oral posttest in the experimental group might have been caused by an improvement in final clusters as students could observe the transcripts of their oral production and realize that a final plural ‘s’ was missing. Students in the control group might produce the final ‘s’ in written production but not being aware of this pronunciation error in the oral posttest. Thus, errors on this level were not significant in the written delayed posttest.

Consequently, these results let us accept the fourth hypothesis (H_4) that grammatical

and lexical improvements are still visible after an extended period.



con. =control, exp.=experimental, s./pl.= singular/plural, s. structure=sentence structure, miss.=missing words

Figure 7: Results of all three tests

(2) 教師教學反思 Reflection on teaching

The error interaction provided in ERLE underpins the main hypotheses being tested out in this experiment. Namely, that use of ERLE can reduce the frequency of errors in students' spoken production. By providing this form of feedback and interaction soon after a sentence was submitted, it enabled students to quickly grasp their common grammatical and lexical mistakes. This raised their awareness of the gap between the language they were producing and the target language and was reflected in students' answers to the questionnaire. The error interaction provided an immediate opportunity to try out new hypotheses on how to fix the sentence submitted. As every single error elicited the error signal, students could be confident that the feedback they were receiving was accurate. It also pushed them to create well-formed sentences, as Saoirse would only continue the conversation and speak back to them after a well-formed sentence was submitted.

Furthermore, the use of ERLE proved to be more effective in reducing pronunciation errors on the phonemes and cluster level compared to PPR, while PPR was more effective on the sentence level. This makes sense, as ERLE gives immediate feedback so that students not only get aware of problems but can also try to immediately remedy them during classes with Saoirse. On the other hand, while the ASR will detect missing or added syllables to some extent, it is unable to detect problems with word stress or problems on the sentence level such as intonation or linking. Feedback by a real person is in this case more effective, even if given delayed as in PPR. This is supported by similar improvements on the word and sentence levels in the control group compared to the ERLE group. Improvements in these two levels in the ERLE group may well have been influenced to a greater extent by classroom instruction than by feedback through the avatar. Classroom instruction alone, on the other hand seems to be more of a waste of time in regard to improving phoneme and cluster problems.

Based on the transcriptions of pre- and posttest audio recordings integrated into the ERLE platform and provided by UCD, use of the ERLE platform also led to a significant reduction in the frequency of hesitations in the spoken language production of our students in comparison to a control group who did not use the platform. Thus, regular conversation classes in ERLE improved students' speech fluency.

A similar positive effect for grammatical and lexical errors was found based on the transcriptions of the integrated pre- and posttests in ERLE and the writing proficiency test at the end of the second semester. As the delayed post-test took place 6 months after the end of the intervention, this shows that improvements gained through the conversation classes in grammar and lexis are long-lasting.

(3) 學生學習回饋 Student feedback

Results of the questionnaire, including means, standard deviation and number of students voting for a specific degree are shown in Appendix (A).

28 questionnaires were completed by the experimental group in the days following the final class with Saoirse and the posttest. Bar charts for responses to the initial statements on Performance Expectancy (PE), Effort Expectancy (EE), Hedonic Motivation (HM) and Behavioral Intention (BI) are shown in Figure 4 with responses varying from 1 ("strongly disagree") to 5 ("strongly agree"). They provide a general overview of the responses by the participants, who gave largely positive feedback on the usefulness, ease of learning and enjoyment of the platform. Intention to continue using ERLE in the future was not as high as the other items, but still one that indicated a positive result.

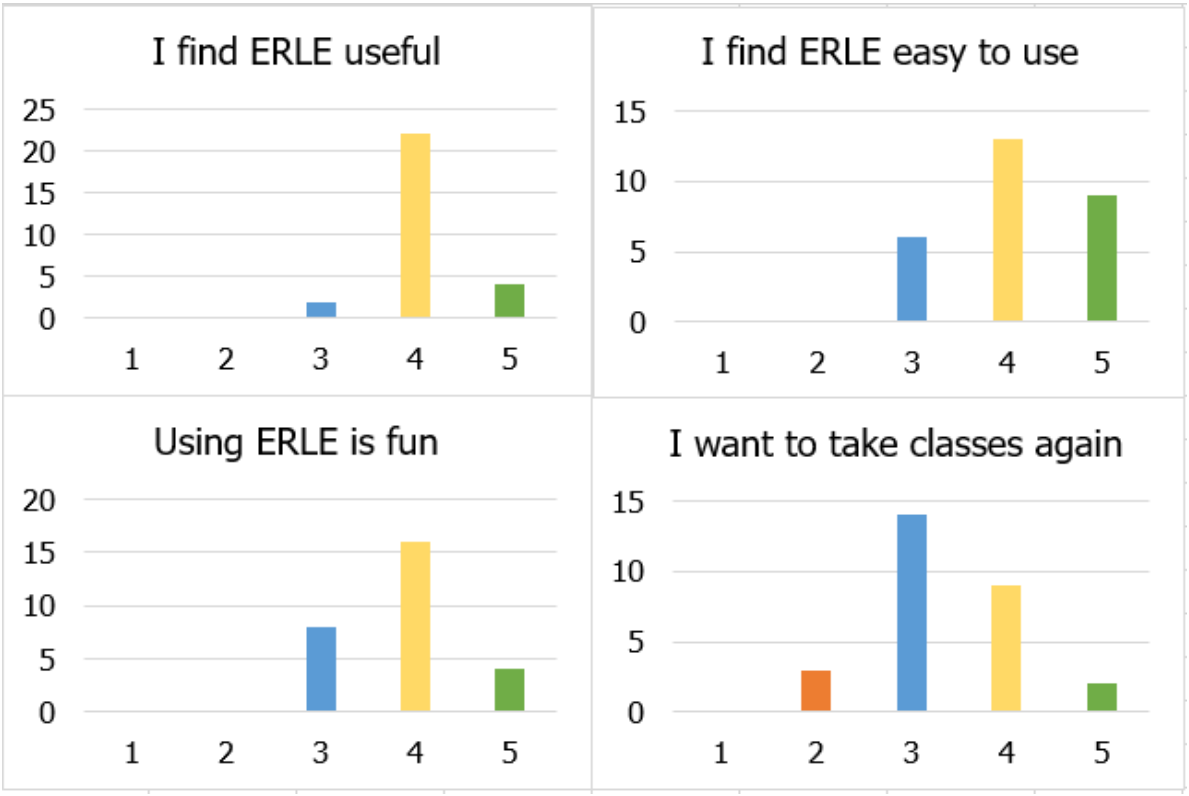


Figure 8: Responses to Primary Questions for PE, EE, HM and BI

When the platform was still in the early developing phase and tested by my students in

the semester preceding this research, issues with response times were one of the most common complaints by the students. Due to this feedback, the platform underwent further development and response time had been improved. Most students (17) reported that they had to wait up to 10 seconds for the avatar to think through their sentences and give a response with a mean response time of 9.43 seconds (SD = 5.75). Only two students reported waiting times of more than 16 seconds. On the other hand, 18 students described the speed of the conversation as neither slow nor fast (see Appendix (A)). Some students remarked that they understood that due to their long sentences or number of errors, Saoirse needed more time for thinking.

When words understood by the ASR were different from the intended words, only few students (2) blamed the computer for this because it was not good at understanding English (mean 2.54 SD 0.79). The majority felt that these errors helped them to realize their pronunciation problems (mean 3.82 SD 0.72) and that they wanted to improve their pronunciation and try again (mean 4.11 SD 0.83). Similar results were reported for grammar problems (I improve after seeing my mistakes: mean 4.18 SD 0.55). While students disagreed with the statement that they prefer talking to Saoirse instead of a real person (mean 2.64 SD 1.28), they still recognized that ERLE was useful for improving their pronunciation (mean 3.89 SD 0.69) and grammar (mean 3.86 SD 0.65). Furthermore, most students preferred taking classes once a week (mean 2.07 SD 0.60) when able to choose between less than a week (1), once a week (2), twice a week (3), or more often (4). Thus, the format of one class per week seems to have suited students well.

6. 建議與省思(Recommendations and Reflections)

It is hoped that the financial problems UCD encountered because of the COVID crisis will soon be overcome so that this project will go on in the near future. On the other hand, we also need to teach our students to embrace technology for learning. We had hoped for larger groups but only one-half of the student population of our freshmen students was willing to take part in the experiment. Most students still see mobile phones and computers more as a tool for social interaction and entertainment. They still depend too much on less effective learning techniques such as rereading textbooks, going over answers, etc. Tools like ERLE on the other hand can combine learning with social interactions, thus leading students to using technology for learning. One of the coming projects at UCD is to develop ERLE further to group communications. Thus, students will be able to take classes with a friend. This might help to alleviate the problem some students mentioned that they sometimes did not know what to talk about during classes with Saoirse.

二. 參考文獻(References)

Arnold, J. (2011). Attention to Affect in Language Learning. *Anglistik - International Journal of English Studies, 22*(1), 11-22.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in second language acquisition, 15*(3), 357-386.

Chang, C.N. & Tsai, W.T (2007). The common grammatical errors Taiwanese students often make in English research *教育科學期刊, 7*(1),1-14.

Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. Cambridge University Press.

Chen, Y. (2002). The problems of university EFL writing in Taiwan. *The Korea TESOL Journal, 5*, 59-79.

Gardner, R.C., & MacIntyre, P.D. (1993). A student's contributions to second-language learning. Part II: Affective variables. *Language Teaching, 26*(1), 1-11. https://doi.org/10.1017/S0261444800000045

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309-365.

Luo, B. (2014). Peer reviewed pronunciation reading. In: Ho W.C. (ed.), *Teaching English Language and Literature at Home and Abroad.* Taichung: Feng Chia University, pp. 99-130.

Luo, B. (2016). Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction. *Computer Assisted Language Learning, 29*(3), 451-476. https://doi.org/10.1080/09588221.2014.963123

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language teaching, 46*(1), 1-40.

MacIntyre, P.D. & Gardner, R.C. (1989). Anxiety and second language learning: toward a theoretical clarification. *Language Learning, 39*, 251-275.

Mackey, A., & Goo, J. (2007). Interaction research in SLA: A meta-analysis and research synthesis. In: Mackey, A. (Ed.), *Conversational interaction in second language acquisition: a series of empirical studies*, Oxford University Press, pp.

407-453.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. *Proceedings of the SIGCHI conference on human factors in computing systems* Ft. Lauderdale, FL: ACM, pp. 249-256.

Norman, D. A. (1988). *The psychology of everyday things* (Vol. 5). New York: Basic books.

Oviatt, S. (2006, October). Human-centered design meets cognitive load theory: designing interfaces that help people think. *Proceedings of the 14th ACM international conference on Multimedia*. Santa Barbara, CA: ACM pp. 871-880.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics. 10*, 209-231

Sloan, J., & Carson-Berndsen, J. (2017, August). Was it something I said? Facial Expressions in Language Learning. *Seventh ISCA Workshop on Speech and Language Technology in Education*, Stockholm: KTH Royal Institute of Technology, pp. 1-6

Sloan, J., & Carson-Berndsen, J. (2018, July). Expressive Data: A student corpus with emotion. *XIX Computer Assisted Language Learning Research Conference – Proceedings*. Bruges: KU Leuven, pp 312-321.

Sloan, J., Luo, B. & Carson-Berndsen, J. (2019, July). Emotional response language education: a first 'in-the-wild' evaluation. In: *Social CALL - XXth International CALL Research Conference – Proceedings*. Hong Kong: Education University of Hong Kong, pp. 240-247

Tan, W. S., Liu, D., & Bishu, R. (2009). Web evaluation: Heuristic evaluation vs. user testing. *International Journal of Industrial Ergonomics, 39*(4), 621-627.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning, 46*(2), 327-369.

Wang, W., & Loewen, S. (2016). Nonverbal behavior and corrective feedback in nine ESL university-level classrooms. *Language Teaching Research, 20*(4), 459-478.

Woodrow, L. (2006). Anxiety and speaking English as a second language. *Regional Language Centre Journal, 37*, 308-328

三. 附件(Appendix)

(A): Answers from the questionnaire: 5=strongly agree, 4=agree, 3=neutral, 2= disagree, 1=strongly disagree
number of classes per week*: >2x=three times or more, 2x=twice, 1x=once, <1x=less than once

Overall	mean	SD	5	4	3	2	1
[ERLE is useful for improving pronunciation]	3.89	0.69	5	15	8	0	0
[My pronunciation has improved after using ERLE]	3.79	0.83	6	11	10	1	0
[ERLE is useful for improving grammar]	3.86	0.65	4	16	8	0	0
[My grammar has improved after using ERLE]	3.68	0.61	2	15	11	0	0
[I find ERLE useful]	4.07	0.47	4	22	2	0	0
[Using ERLE helps me improve my English ability]	4.07	0.60	6	18	4	0	0
[If I continue to use ERLE, my English will improve]	4.00	0.54	4	20	4	0	0
[Using ERLE is fun]	3.86	0.65	4	16	8	0	0
[I would recommend ERLE to my friends]	3.89	0.83		17	11	0	0
Effort							
[Learning how to use ERLE was easy for me]	4.14	0.71	9	14	5	0	0
[My interaction with ERLE was clear and understandable]	4.04	0.69	7	15	6	0	0
[I find ERLE easy to use]	4.11	0.74	9	13	6	0	0
[It was easy for me to become skillful at using ERLE]	4.04	0.64	6	17	5	0	0
Conditions							
[I have the resources necessary to use ERLE]	3.61	0.79	3	13	10	2	0
[I have the knowledge necessary to use ERLE]	3.64	0.78	4	11	12	1	0
[ERLE is compatible with other technologies I use]	3.79	0.74	5	12	11	0	0
[I can get help when I have difficulties using ERLE]	3.96	0.92	8	13	6	0	1
[The website loads quickly]	3.07	0.94	1	9	10	7	1
[After recording, the words are shown to me quickly]	3.11	0.92	0	12	8	7	1
[After submitting a sentence, Saoirse responds quickly]	2.96	0.92	1	8	8	11	0
[After making a mistake, I get feedback quickly]	3.68	0.86	5	11	10	2	0
[Internet problems happen frequently]	3.21	1.07	3	9	8	7	1
After recording my voice...	mean	SD	always	often	sometimes	seldom	never
[the words I saw were different from what I said]	3.14	0.65	0	8	16	4	0
[I tried to fix the wrong words by speaking again]	3.57	0.84	4	10	12	2	0
[I fixed the wrong words by tapping and typing]	3.50	1.11	5	10	9	2	2

When the words I see are different from what I say...	mean	SD	5	4	3	2	1
[I feel frustrated]	3.14	0.93	2	7	13	5	1
[I think the computer is not good at understanding English]	2.54	0.79	0	2	14	9	3
[I can realize my pronunciation problems]	3.82	0.72	4	16	7	1	0
[I want to improve my pronunciation and try again]	4.11	0.83	10	12	5	1	0
During class							
[I learn a lot from seeing my own mistakes]	4.14	0.65	8	16	4	0	0
[I make more mistakes than I previously thought]	2.89	0.99	0	10	7	9	2
[I make the same mistakes repeatedly]	2.71	0.81	0	5	11	11	1
[I improve after seeing my mistakes]	4.18	0.55	7	19	2	0	0
[Talking to Saoirse is similar to talking to a real person]	3.07	1.02	1	10	9	6	2
[I prefer talking to Saoirse instead of a real person]	2.64	1.28	2	5	10	3	8
[Saoirse's expressions made me interested in using ERLE]	3.36	0.87	2	11	10	5	0
[Saoirse's expressions were natural]	3.04	1.00	3	4	13	7	1
[Saoirse's expressions were strange]	2.93	1.02	1	8	9	8	2
When Saoirse looks away after I make a mistake							
[I feel frustrated because I was wrong]	2.89	0.96	2	3	15	6	2
[I feel happy because I can learn about my mistakes]	3.96	0.79	7	14	6	1	0
[I want to try again to fix my own mistake without help]	3.54	1.32	8	8	6	3	3
[I want to see the correction without trying again]	3.21	1.32	6	6	7	6	3
After class...							
[I like to read my sentences]	3.36	0.83	0	10	15	3	0
[I listen again to my own recordings]	2.71	1.01	0	6	10	12	0
[Reviewing recordings helps improve my pronunciation]	3.64	0.68	0	17	10	1	0
[Reviewing sentences helps improve my grammar]	3.79	0.69	0	18	10	0	0
Future Classes							0
[I want to take classes again on ERLE in the future]	3.36	0.78	0	11	14	3	0
[I plan to continue to use ERLE frequently]	3.00	1.12	0	7	16	3	0

Number of classes/week*	mean	SD	>2x	2x	1x	<1x
[I want to take classes on ERLE]	2.07	0.60	0	6	18	4

Saoirse needed	mean	SD	1-5	6-10	11-15	15-20	20-25
[X seconds to think and respond?]	9.43	5.75	9	10	7	1	1
			slow	neutral	fast		
[The speed of the conversation was]	1.79	0.57	8	18	2		

Please give your opinion on how useful ERLE is for practicing and improving your pronunciation:

It always find my wrong pronunciation even it is smallest problems

It's nice to have such time to talk with someone in English and it really helps me a lot, when facing errors that I made I will think about it and after being corrected I will try to say the sentence again, this helps me a lot though sometimes I will forgot and make the same mistakes again but at least I got someone to correct me and make me thinking about what should I talk to her every week.

I know how to pronounce in the correct way

You can able to see what sort of mistakes you did

I know my pronunciation problems from the words appear on screen, so I can know that I should practice which pronunciation of the word.

It's very useful to me because I don't have many chances to speak English at school or in my daily life.

It can realize my pronunciation mistake, so I can try again and correct my mistakes

I think it's not as helpful as one might need it to be. It gives us the opportunity to speak English more, so it definitely helps us with our speaking. However, I don't think I improved a lot of my pronunciation skill. I guess it has something to do with the time limit. Since I have to speak and think at the same time, 15 seconds [for one sentence] is not enough for me to take my time with my pronunciation.

I think it's more useful for me to improve my grammar

I improve my conversation skill every time I use ERLE

It helps you to practice your pronunciation

It is useful for the people who wanna improve their speaking ability.

I can find out that the pronunciation of my vowels are wrong

It will show the words I speak and catch my wrong words to make me revise.

I could learn from my mistakes. For example, I've never realized the word that I was saying incorrect all the time up until I started to use ERLE.

Having a chance to have a English conversation

ERLE gave me a chance to open my mouth and practice to improve my English speaking skill.

Please give your opinion on how useful ERLE is for practicing and improving your grammar

My mistakes were mainly on grammar and that's why I started this course. And the course was once a week so sometimes I will made the same mistakes for I forgot.

I know how to make my sentences look better

The teacher will help me revise the sentence, and then repeat the sentence to deepen the impression

It shows you the correct sentences

When I make mistakes on ERLE, I will think and try it again. Sometimes, I don't know where is the mistakes, ERLE will correct it, so I can learn the mistakes and grammar.

It can tell me where is my mistakes.

I can quickly notice my grammar mistake.

I think that I learned and improved my grammar more than my pronunciation.

It will correct my wrong grammar

It helps me to realize my grammar mistakes

It tells you where is wrong

When I make sentences, I need to speaking correctly or it won't give me points.

It's just like "Grammarly" I think it's good

When my words are sent to Saoirse, if my words have problems Saoirse will be confused so that I can know my mistakes.

it was giving correction. therefore it was useful

Cant speak with concentrate on grammar

How did you feel when seeing Saoirse's expressions change?

Feel like Saoirse is a real person(i know she really is :))

Interesting, I'm so curious to see her reaction every time I finish my sentence.

Weird but it's not that much

I feel very funny

Weird smile with front teeth showed

I will be curious about what wrong with my recording.

Cool~

I feel interested in talking to her because I think she is listening to me patiently.

I felt nothing.

Good, that's interesting and let me feel less nervous.

Interesting

Nothing

Just like a normal person

Well not too good. Because I know she's not a real person

I feel very interesting.

Felt good.

Only a little bit to feel it is real

What would you improve about the avatar?

When I finished my sentence it took a long while for her to think and we only got thirty minutes to talk so if she could think faster maybe we can talk more and practice more.

Maybe I would make her look cuter

I don't think it needs to be changed

By recreate a new avatar

The intonation can be more human nature, such as happy intention and sad intention.

Make it looks more natural.

Sometimes she cant hear me even i have repeated for three times. I hope I can correct my word by typing at least two words.

I think it will be fun if we can chose the appearance and the name of it.

Maybe can choose a girl or a boy?

None

It looks weird

A lot

By trying to make Saoirse know what I mean.

I would improve avatar's speed of writing

Make it more like real person

Additional comments

Well done

Thank you for giving me this opportunity to use ERLE. Facing her can make me speak English less nervously. Although I have a lot of problems with pronunciation, the teacher will tell me where I went wrong. To improve my pronunciation.

Hope that the radio frequency reduce to minimum

I think you can give more detailed instructions or explain what functions I can use at the beginning. For example, how to correct the wrong words.

I hope the nation when signing up can add Hong Kong.

Hope ERLE be better and be used by the people all over the world XD

I hope I can improve my English speaking soon. What's more, the ways to improve my English pronunciation on ERLE can be more interesting so that I can use it happily

the format sometimes isn't correct

It works pretty well

Again I hope the speaking time can be free and I don't like the feeling of talking with machine. But ERLE is really helpful to my English study. I hope I can take more classes in the future.

I think the internet is the most worst thing , each week i have received the bad internet news , it always affect me to talk with the teacher

actually I think talking to real person is effective but the ERLE is funny for its avatar

It would be nice if ERLE could become a download app.

ERLE makes me better organized when I speak English

(B): Intensity of pronunciation problems

phoneme level		pretest	posttest	difference	significance (p)
ERLE	mean	2.57	1.59	-0.98	0.000
	SD	0.47	0.40	0.24	
PPR	mean	2.30	1.82	-0.43	0.000
	SD	0.68	0.65	0.29	
control	mean	2.46	2.36	-0.10	0.284
	SD	0.68	0.64	0.46	
significance	(p)	0.346	0.000	0.000	
cluster level		pretest	posttest	difference	significance (p)
ERLE	mean	2.11	1.52	-0.59	0.000
	SD	0.79	0.63	0.40	
PPR	mean	2.05	1.75	-0.30	0.000
	SD	0.77	0.69	0.37	
control	mean	2.16	2.16	0.00	
	SD	0.47	0.41	0.41	1.000
significance	(p)	0.847	0.001	0.000	
word level		pretest	posttest	difference	significance (p)
ERLE	mean	1.25	0.64	-0.61	0.000
	SD	0.70	0.56	0.63	
PPR	mean	1.23	0.73	-0.50	0.002
	SD	0.61	0.55	0.66	
control	mean	1.12	0.74	-0.38	
	SD	0.73	0.50	0.58	0.003
significance	(p)	0.783	0.778	0.443	
sentence level		pretest	posttest	difference	significance (p)
ERLE	mean	2.70	2.11	-0.59	0.000
	SD	0.45	0.46	0.43	
PPR	mean	2.80	1.77	-0.98	0.000
	SD	0.33	0.51	0.52	
control	mean	2.48	1.94	-0.50	0.000
	SD	0.59	0.74	0.50	
significance	(p)	0.070	0.168	0.000	

(C): Intensity of grammar problems in percent of total word count

articles		pretest	posttest	difference	significance (p)
ERLE	mean	1.92	1.47	-0.46	0.078
	SD	1.59	1.26	1.66	
control	mean	1.47	2.08	0.61	0.047
	SD	1.33	0.352	1.77	
significance	(p)	0.132	0.115	0.014	
singular/plural		pretest	posttest	difference	significance (p)
ERLE	mean	2.64	1.16	-1.48	0.002
	SD	2.43	1.31	2.55	
control	mean	2.63	3.45	0.83	0.114
	SD	3.00	3.09	3.40	
significance	(p)	0.493	0.000	0.003	
verb		pretest	posttest	difference	significance (p)
ERLE	mean	1.96	1.96	0.00	0.499
	SD	2.05	1.54	2.73	
control	mean	2.31	2.55	0.24	0.316
	SD	1.79	2.04	2.50	
significance	(p)	0.254	0.008	0.370	
sentence structure		pretest	posttest	difference	significance (p)
ERLE		1.83	0.91	-0.92	0.009
		2.01	0.97	1.91	
control		1.65	1.70	0.05	0.450
		1.67	1.84	2.05	
significance		0.361	0.026	0.039	
lexis		pretest	posttest	difference	significance (p)
ERLE	mean	2.84	2.69	-0.15	0.410
	SD	2.88	2.20	3.36	
control	mean	2.62	3.85	1.23	0.027
	SD	2.20	2.83	3.11	
significance	(p)	0.380	0.049	0.063	

(C): Intensity of grammar problems in percent of total word count (continued)

missing words		pretest	posttest	difference	significance (p)
ERLE	mean	2.27	2.13	-0.14	0.344
	SD	1.67	1.87	1.79	
control	mean	1.64	2.98	1.34	0.019
	SD	1.54	3.14	3.10	
significance	(p)	0.156	0.116	0.018	
total errors		pretest	posttest	difference	significance (p)
ERLE	mean	15.86	12.12	-3.74	0.001
	SD	6.41	5.52	6.03	
control	mean	14.60	19.25	4.65	0.006
	SD	8.48	10.47	8.82	
significance	(p)	0.541	0.001	0.000	