17

Artificial Intelligence Technology for EAP Speaking Skills: Student Perceptions of Opportunities and Challenges

Bin Zou, Sara Liviero, Mengyuan Hao and Chaoyang Wei

Introduction

This study aims to contribute to international research on integrating technology to English for academic purposes (EAP) curriculum design by documenting Years 1 and 2 (Y1 and Y2) EAP students' perceptions of how artificial intelligence (AI) mobile applications (apps) currently available support their preparation for engaging in EAP speaking tasks and assessment at an English-speaking university in China, which is increasingly providing for international students (Park & Slater, 2014). Specifically, student attitude is explored in the AI context because if students become

B. Zou (✉) • S. Liviero • M. Hao
Xi'an Jiaotong-Liverpool University, Suzhou, China
e-mail: Bin.Zou@xjtlu.edu.cn; Mengyuan.Hao16@student.xjtlu.edu.cn

C. Wei
University of Liverpool, Liverpool, UK
e-mail: Chaoyang.Wei16@student.xjtlu.edu.cn

M. R. Freiermuth, N. Zarrinabadi (eds.), *Technology and the Psychology of Second Language Learners and Users*, New Language Learning and Teaching Environments,
https://doi.org/10.1007/978-3-030-34212-8_17

interested in and would like to spend more time working on language learning activities, they tend to have higher motivation towards the learning process in which they are engaged (Gardner & Lambert, 1972). Thus, it is vital to examine student attitude and interest in the computer-assisted language learning (CALL) context especially related to AI. Many studies have investigated student attitude in the CALL environment and found learners' positive attitudes towards using technology in language learning, for example, Kessler, Bikowski and Boggs (2012); Zhang, Song and Huang (2014); Zou, Wang and Xing (2016); Xu and Peng (2017); and Zou, Li and Li (2018). However, there has been scant research investigating students' attitudes about applications designed for mobile devices that are available to them (see Glossary).

China intends to invest significantly in the development of AI tools for educational purposes, aiming to take a leading role in this field (Knight, 2017). Private companies as well as the most prominent Chinese universities have created AI laboratories for the development of educational programmes that aim to provide for the needs of national and international university students. Kim, Soyata and Behnagh (2018, p. 5308) predict that by 2024, 'smart classrooms' will be characterized by a digital-system environment providing teachers with real-time presentation feedback on the effectiveness of their communicative skills. Meanwhile, research studies have noted how the teaching of EAP speaking skills still largely relies on teacher input, marking and feedback, and it is mainly limited to the teaching of suprasegmental EAP speaking skills. This often involves instructions to deliver presentations and spoken texts as well as requiring self-study tasks (Gilakjani, 2011; Gilakjani & Sabouri, 2016) on their computers and/or mobile devices (Reinders & Darasawang, 2012). Although these devices are favoured by students because of their affordances of time and space (Al-Fahad, 2009; Demouy & Kukulska-Hulme, 2010), they provide neither detailed, reliable feedback on speaking performance related to EAP speaking exam descriptors (Wang & Young, 2014) nor real-time, authentic interaction typical of face-to-face communication (Bordonaro, 2003).

In the Chinese university where this study was based, EAP provision consists of a course covered in the first two years—four semesters comprehensively—of undergraduate studies. Mainly, the undergraduate student population consists of speakers of Chinese as a first language

(CL1); however, the university aspires to consolidate its international status and is endeavouring to attract an increasing number of students from various linguistic backgrounds, benchmarking their English proficiency on levels of the International English Language Testing System (IELTS, 2018). EAP courses are content-based and do not focus on isolated linguistic skills; EAP summative assessment components, such as discrete speaking and writing skills are focused on instead, which are complemented by an integrated exam assessing listening, reading and writing skills. The typical tasks of Years 1 and 2 EAP formative and summative assessments of speaking skills are as follows:

- Academic presentation of a generic topic covered during the course
- Academic presentation from the students' subject area
- Small-group discussion on a generic course topic
- Answering direct subject-specific and generic course topic questions

Performance on the above tasks is judged on specific EAP speaking descriptors, inspired by those outlined by the Common European Framework of References for Languages (CEFR) (Council of Europe, 2018; de Jong & Benigno, 2018) for language learning, teaching and assessment, and the BALEAP (2018) framework for EAP syllabus design and assessment. These include:

- Clarity of pronunciation, effective stress and intonation, maintained fluency (rhythm and pace)
- Subject-specific vocabulary and grammatical accuracy
- Development of relevant, logically organized ideas including details and examples
- A good range of discourse markers and functional language
- Demonstration of the ability to interact with other speakers

Increasingly, there has been an acknowledgement that EAP curricula have come under the influence of the internationalization agenda essentially supporting the notion of English as a lingua franca (ELF) (Jenkins, 2014, 2017); however, this idea is still characterized by non-native-like English pronunciation (Kang, Thomson, & Moran, 2018; Kim, 2006) making it widely unrecognized and unappreciated by most English

Language Learning (ELL) computer application developers, (Chiu, Liou, & Yeh, 2007; McCrocklin, 2016), thus increasing EAP student dependence on face-to-face interaction when practising English-speaking skills (Bordonaro, 2003; McCrocklin, 2016).

This project wishes to document research and practical gaps resulting from AI-assisted EAP speaking skills, an underdeveloped and underresearched line of inquiry when considering both linguistic and technological perspectives.

Literature Review

Current Approaches to EAP Speaking Skills

Students acquire speaking skills to address specific tasks if they are systematically taught to rehearse and put into practice tasks such as EAP presentations (Bruce, 2011). However, without feedback, students struggle to identify either strengths or weaknesses in their speaking skills, often due to their lack of subject knowledge and/or the pedagogical content knowledge required to conduct an expert analysis (Celce-Murcia, Brinton, Goodwin, & Griner, 2010; Dlaska & Krekeler, 2008; McCrocklin, 2016). Gilakjani (2011) identified the need for software that assists generic English language learning to adopt a scaffolded approach with the aim of helping students develop speaking skills autonomously but within parameters that offer activities and feedback related to EAP linguistic and performance goals. Essential segmental and suprasegmental features of speech should be planned, drafted and rehearsed to improve their EAP presentations (Bruce, 2011), pronunciation and consequently the quality of intelligibility, that is, the quality of pronunciation that ensures intelligibility for academic presentations (Setter & Jenkins, 2005; Wang & Young, 2014). However, it is crucially important to progressively integrate non-native EAP models of pronunciation into EAP teaching and assessment material, in order not to alienate students' perceptions of self and of the prestige and knowledge typical of international university research and teaching staff (Murphy, 2014).

Recently, researchers have been contemplating the possibilities to train technology to recognize non-native accents (NNAs), thus elevating the functionality of computer applications to accurately represent international academic students and staff from non-native English speaking backgrounds (Beaven & Neuhoff, 2012; Mauranen, 2012).

The available AI-ELL applications (from this point forwards simply 'apps') mainly templates human-computer interaction of 'native English' phonetic representations (Hincks, 2005), and as such, a reasonable question to ask is whether this affects the degree of accuracy AI-ELL app feedback achieves to score EAP student performances, which may be accurately structured in spoken EAP texts but which are not closely reflected in the range of native English-speaking pronunciation that is predominantly featured in high-stakes test materials (TOEFL-iBT, IELTS, TOEIC). In essence, this promotes deficit models of non-native English accents (Kang et al., 2018, p. 3; Murphy, 2014).

In light of this, areas that AI-EAP should be concerned with when developing applications that are useful for assisting EAP speaking skills are (a) training AI-EAP to recognize accurate, intelligible EAP pronunciation; (b) afford assessment-related task practice and feedback; (c) train teachers to incorporate AI-EAP tools in syllabus design according to the scaffolded approaches identified; and (d) train students to use AI-EAP autonomously.

AI and VRT: Potential and Limitations

The progress of voice recognition technology (VRT) in CALL has given great impulse to the creation of apps assisting the learning of speaking skills, in addition to the prevalent attention for listening, reading and writing (Kim, 2006). There are various mobile applications and computer programmes featuring VRT, such as *Nuance Dragon Dictation* (2018), *Praat* (Boersma & Weenink, 2018) and *Duolinguo* (2018), available for free, easily downloadable on most platforms and that can be used both online and offline (Meisam & Tavakoli, 2015).

The positive effects of VRT introduction in the teaching of English pronunciation were reported by Derwing, Munro and Carbonaro (2012), who also included a review of VRT limitations in recognizing NNA

(Non-Native Accents) (also see Liakin, Cardoso & Liakina, 2015). In addition, Deng and Training (2015) reviewed apps focused on both intentional and accidental vocabulary learning strategies. Poignant to the present study, real-time feedback of student speaking assessment performance has long been a research and practical concern (Bernstein, Cohen, Murveit, Rtischev, & Weintraub, 1990), despite reports that such feedback may prove inhibiting to Chinese learners of English (Jia, 2009). Existing applications currently used by students are reportedly offering limited performance feedback and evaluation due to limited prosodic system features that result in generic guidelines on student pronunciation that would not fit the purposes of EAP speaking skills in higher education.

To investigate how apps could be integrated into EAP courses, Zou, Li and Li (2018) selected 84 students from four classes of majors in economics and marketing subjects at a university in China, taking EAP courses with English CEFR levels of proficiency between B1 and B2 (Council of Europe, 2018). The majority of participants held positive attitudes towards mobile learning in class because of the convenience of access to course-relevant materials and stimulating student-teacher interaction. Kan and Tang (2018) also found students' positive attitudes towards using apps for speaking practice. Negative perceptions of CALL encompassed discomfort caused by reading from electronic screens, and by limited Wi-Fi access, especially when the apps required good network connections (Al-Fahad, 2009; Liakin et al., 2015). Negative perceptions of VRT are focused on practical and technical aspects, such as the associated high costs of CALL digital technology, of its maintenance and apps' limited accessibility when needing complicated instructions (Chen, 2011; Meisam & Tavakoli, 2015).

A priority for research (Chavan & Gawande, 2015) has been to focus on enhancing the quality of VRT when speakers engage in a conversation or give presentations. As Douma, Anderson, Akahane and Mizikovsky (1996) identified, current voice recognition is capable of attaining accuracy at 90% when recognizing native accents; however, it performs less well with non-native accents (Neri, Cucchiarini, & Strik, 2003). An increasing number of students in Chinese universities have been reported to be using mobile apps to improve English language skills during lessons and independently (Zou, Li, & Li, 2018).

As learning a foreign language for academic purposes requires a higher degree of proficiency, CALL software should be designed with the specific purpose of enabling students' independent access to feedback on their performance (Köse & Arslan, 2014). These perspectives highlight the important role that AI might play in assisting EAP teaching and learning.

Mainstream VRT-Assisted AI-ELLs Apps

There are some VRT products in the market; for example, Siri (Apple Inc., 2018) was created in 2011 as a virtual user assistance for users of Apple devices such as iPad and iPhone. Moreover, Google Cloud Speech-to-Text (Google Cloud, 2018) transforms audio to text by means of an application programme interface (API) that can recognize a number of languages. Functions include voice command and control, voice to text transcription and live streaming.

In China, three main AI-ELL apps have been developed for the Chinese markets; they are *Chivox, Ltd*. (2018a), *iFlytek* and *Liulishuo*. For each product, the following descriptions and evaluations of their English language learning-assisting features will be discussed.

First, Chivox technology (Chivox, Ltd., 2018b) claims to assist the English language learning abilities regarding (a) pronunciation, (b) task-based conversation and (c) presentation. Chivox provides exercises reinforcing and assessing clear pronunciation, fluency, stress and intonation of morphemes; words and sentences; and paragraph as well as text formation. It can be used by students as an independent learning tool. A feature of Chivox is 'Kami English,' a resource for oral English practice that claims to use intelligent speech analysis technology to cater to K12 curricula (Chivox, Ltd., 2018a; K12 Inc., 2018).

iFlytek (2018) developed AI products focused on English language learning for generic as well as other purposes, such as medical English. In the research field of intelligent VRT, iFlytek claims to be committed to developing technology for artificial understanding and evaluation of natural language and handwriting. iFlytek uses voice recognition to identify users' identity through the recognition of speech signals. Among its various app products, the RealSkill app was developed for the evaluation of

TOEFL and IELTS speaking and writing. For a given topic of TOEFL, it records no more than 45 seconds of voice input and offers a numeric score according to topic development, delivery and language use.

Liulishuo (2017) is another tool applying AI technology for Chinese students to practise English speaking. Specifically, there is a version of Liulishuo designed completely for IELTS tests, and this IELTS version will be discussed in detail in this section. First, it provides free model tests with virtual scenes and examiners to recreate the real speaking exam. The whole test procedure is similar to the real test, lasting around 15 minutes, and all voice responses are recorded. Subsequently, it takes 30 minutes to evaluate the performance and release detailed feedback. The feedback mainly covers fluency, vocabulary, grammar and pronunciation making it quite similar to RealSkill; however, Liulishuo for IELTS might offer a more detailed assessment by comparison. First, the speaker's willingness to communicate and an answer's relevance is commented on in the fluency evaluation. Second, overused words in the responses are listed and substitute words are suggested in the meantime, which tends to help users broaden their lexical range. Subsequently, quantifier mistakes are detected in the grammar evaluation. Overall, the feedback of Liulishuo for IELTS is considered helpful. Most importantly, Liulishuo can record up to two minutes of voice input, while the maximum for RealSkill is 45 seconds. Although its advantages are many, it has some of the same problems as RealSkill. First, it is also difficult for the system to distinguish every utterance perfectly. In other words, it cannot recognize every word with 100% accuracy. Second, although it intends to recreate the authentic exam environment, the app cannot replace a real examiner's vivid interaction with students.

AI applications such as those mentioned here are only as good as their perceived usefulness by students. This study, then, looks at students' attitudes towards using such applications outside of the language classroom.

Methodology

A mixed methods approach was adopted, making use of a student questionnaire and semi-structured interviews, whereby qualitative results assisted in explaining and interpreting the findings of the questionnaire

(Dörnyei, 2007). The adopted sequential explanatory mixed methods design consisted of a survey of student perceptions of current use and efficacy of AI-ELL apps for speaking skills, followed by qualitative interview data collection and analysis (Creswell, 2013) to explain and interpret the questionnaire's findings, thus gathering rich data on students' perceptions of the use and efficacy of AI-ELL apps (Dörnyei, 2007). To gain an understanding of students' perceptions while using the applications, the study was guided by the following research questions:

1. What are the opportunities that AI offers in the development of tools for speaking skills in English language learning as based upon students' attitudes ?
2. What are the challenges in developing AI tools for speaking skills in English language learning as based upon students' attitudes ?

Quantitative Data Collection

This small-scale research project aimed to map trends (Oppenheim, 1992, pp. 12–13) of knowledge levels and of attitudes towards AI-ELL apps commonly used by Y1 and Y2 students of EAP courses from several Chinese international universities. Data were collected through the distribution of a questionnaire specifically designed to elicit appropriate responses with respect to the study's aims, further described below.

A non-probability, convenience sample was drawn from Year 1 (Y1) and Year 2 (Y2) EAP students. Completed questionnaires were returned by 113 respondents; however, only 16 were from participants of other Chinese universities. Respondents comprised 34 Y1 and 79 Y2 students. A process of random sampling would not have been possible within the limited time scale of this study (Sapsford, 1999).

The questionnaire used for the quantitative aspect of the study was specifically designed for the purposes of this study, though informed by the research reviewed in terms of relevant instruments previously used; their theoretical and logical guidelines were employed to evince student attitudes and knowledge of AI-ELL apps (Cargill & O'Connor, 2011;

Glasman-Deal, 2010). The questionnaire comprised 23 items and consisted of the following sections:

- Section A: Background information: year of attendance, EAP course level and nationality.
- Section B: Attitudes towards AI-ELL apps/EAP apps.
- Section C: Currently used AI-ELL apps/EAP apps.
- Section D: Students' perceived needs when revising for Y1 and Y2 EAP speaking exams.
- Section E: Students' perceived limitations of AI-ELL apps/EAP apps compared to EAP exam requirements.
- Section F: Consent to be considered for individual interview.

Participant perceptions were scored on a five-point Likert scale (Pallant, 2013) ranging from 'strongly agree' to 'strongly disagree' and 'true or false' type questions. Reliability and validity of the knowledge and attitude test instruments were investigated by utilizing the same data set used to explore the key hypotheses.

Validity and Reliability Report

The reliability and validity of the knowledge and attitude test instruments were investigated by utilizing the same data set used to explore the key hypotheses. A Cronbach $\alpha = 0.750$, $0.7 \leq$ Cronbach $\alpha < 0.9$ indicated adequate reliability (Pallant, 2013). However, a Kaiser-Meyer-Olkin (KMO) value equal to 0.489 (less than 0.5), indicated a potential instrument's poor factor analysis. Future research should ensure that more consistent scales are used for each questionnaire section; that this be piloted again in order to improve its consistency.

Procedure

The questionnaire was first pilot tested with a small number of people of a similar background to the actual research sample, in order for the

Table 17.1 Assessing the normality of data

One-Sample Kolmogorov–Smirnov Test									
		1. Please indicate your grade	[AI-ELL app activities to improve speaking skills] 15. About following features of AI tools for EAP Speaking practice, indicate how important you think they are	10. If you use any, how often do you use AI Tools to practise EAP speaking skills?	11. practising English with AI Tools for EAP Speaking Skills will improve my English skills overall	3. Please indicate your EAP Stream	2. Please indicate your university	22. Do you think AI Tools for EAP speaking skills can replace face-to-face teaching?	[Price (if not affordable or acceptable)] 21. What limitations do you think speaking technology might present? Please rank each possibility
N		113	113	109	113	113	113	112	113
Normal Parameters[a,b]	Mean	1.70	2.71	3.00	3.01	1.51	1.14	1.81	2.78
	Std. deviation	0.461	1.075	0.782	0.940	0.721	0.350	0.392	1.140
Most extreme differences	Absolute	0.442	0.181	0.289	0.248	0.381	0.515	0.496	0.188
	Positive	0.257	0.181	0.243	0.247	0.381	0.515	0.316	0.166
	Negative	−0.442	−0.173	−0.289	−0.248	−0.238	−0.343	−0.496	−0.188
Test statistic		0.442	0.181	0.289	0.248	0.381	0.515	0.496	0.188

(*continued*)

Table 17.1 (continued)

One-Sample Kolmogorov–Smirnov Test								
	1. Please indicate your grade	[AI-ELL app activities to improve speaking skills] 15. About following features of AI tools for EAP Speaking practice, indicate how important you think they are	10. If you use any, how often do you use AI Tools to practise EAP speaking skills?	11. practising English with AI Tools for EAP Speaking Skills will improve my English skills overall	3. Please indicate your EAP Stream	2. Please indicate your university	22. Do you think AI Tools for EAP speaking skills can replace face-to-face teaching?	[Price (if not affordable or acceptable)] 21. What limitations do you think speaking technology might present? Please rank each possibility
Asymp. Sig. (2-tailed)	0.000[c]	0.000[c]	0.000[c]	0.000[c]	0.000[c]	0.000[c]	0.000[c]	0.000[c]

[a]Test distribution is normal
[b]Calculated from data
[c]Lilliefors significance correction

appropriate changes to be made (Oppenheim, 1992, pp. 128–30). The final questionnaire was submitted online and returned completed by 113 Y1 and Y2 students. Subsequently, six participant students were selected for semi-structured interviews on their perceived needs in using AI-ELL apps for the development of their EAP speaking skills.

The items on the questionnaire focused on exploring student perceptions of their experience when using AI for speaking practice. Survey data analysis informed the formulation of the semi-structured interview, aimed to gather in-depth perspectives on concepts that emerged from the questionnaire analysis.

The hypotheses were tested by using parametric and non-parametric statistical tests carefully selected for each of them in relation to individual tests' requirements using the Statistical Package for the Social Sciences (IBM SPSS, 2018).

Background and Characteristics of the Sample

The majority of respondents, 69.9%, were Y2; the remaining 30.1% were Y1. While 85.8% of respondents were students studying at the project university, only 14.2% were from other Chinese universities, and only one was attending a university outside China. Most respondents were EAP main stream course students and only 24.8% reported to attend high-level courses; and 13.3% of respondents, corresponding to the percentage of students from external universities, indicated not to know what EAP meant. The survey revealed that 3.5% of respondents were Indonesian and 96.5% of respondents were Chinese.

Assessing the Normality of Data

The normality of data was assessed by the Kolmogorov-Smirnov statistical test (Pallant, 2013). Table 17.1 illustrates that the result obtained was $p = 0.000$ for all variables. A result of $p < 0.05$ indicated significance, probably due to the need to further adjust measuring scales to each questionnaire item. Results therefore should be considered cautiously;

nevertheless, while highlighting the time and resource constraints of its procedures, the study is also aware of the attention it brought to this under-researched field and the potential it offers in the refining of its tools for further research.

Qualitative Data Collection: Sample Selection Criteria

Six questionnaire participant students were selected for the semi-structured interview on their perceptions of the opportunities and limitations in using AI tools for the development of their EAP speaking skills. Although survey data analysis informed the formulation of the semi-structured interview along themes outlined in 3.1 (paragraph 2), the semi-structured nature of the interview tool allowed new themes to emerge from individual participants and thus gather in-depth perspectives on concepts that emerged from the questionnaire analysis (Charmaz, 2003).

Data Analysis

Semi-structured interviews aimed to probe the questionnaire results and were analysed inductively (Thomas, 2006). In particular, interviews further investigated the following:

1. Students' knowledge of current AI-ELL app tools
2. Frequency of AI-ELL app use
3. Students' perceived difficulties in EAP speaking and learning
4. Students' perceived advantages of AI-ELL apps /AI-EAP apps
5. Students' perceived limitations of AI-ELL apps /AI-EAP apps

Interviews were transcribed and the analysis proceeded by applying initial coding to all participant answers in the same order as the interview questions. A second phase of coding identified relevant emerging topics, in turn, compared with data emerging from the questionnaire, and a third phase of selective coding yielded the themes illustrated in the following sections. Student interviewees were coded as S1 (student 1, 2, etc.) and gender was indicated as M/F (male/female).

Findings

Quantitative Data Analysis: H1: Year of Attendance and Attitudes About AI-ELL Apps for Speaking Skills

In testing attitudes, participants, using a five-point Likert scale ranging from 'very important' to 'not important,' gave scores as a way to indicate what attitudes they held concerning AI-ELL app activities and exercises aimed to improve speaking skills. Figure 17.1 shows that the mean for the attitude test was M = 2.71(SD = 1.075 N = 113). While 43.4% of participants believed AI-ELL app activities contributed to the development of speaking skills, 21.2% of participants thought that such app activities did not.

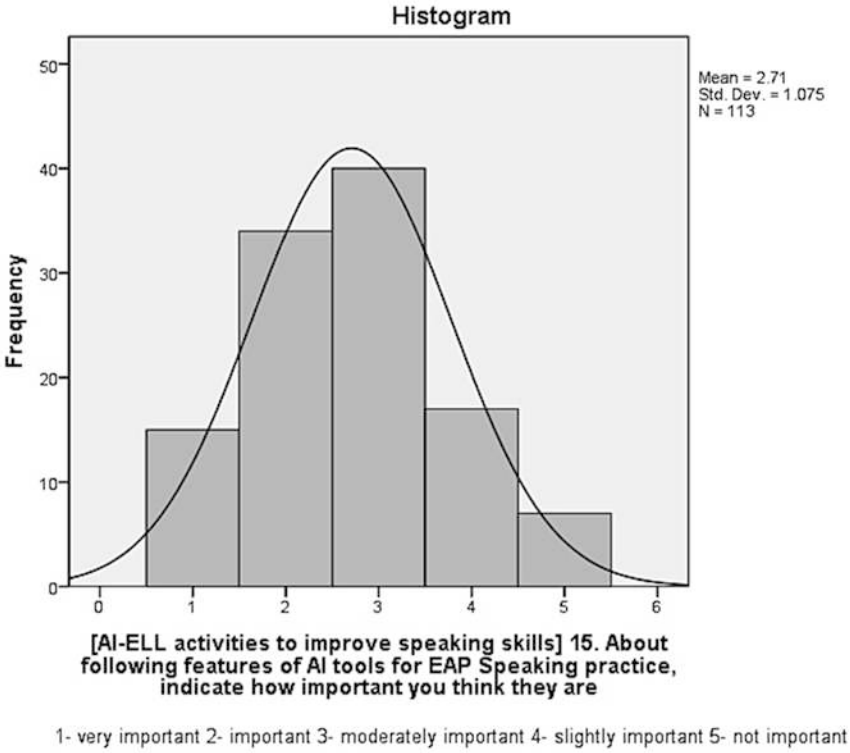

Fig. 17.1 Attitudes and knowledge: AI-ELL apps for speaking skills

Table 17.2 H1 and students' attitudes towards AI-ELL's potential to improve their speaking skills

Independent samples test										
		Levene's test for equality of variances		*t*-Test for equality of means						
									95% confidence interval of the difference	
		F	Sig.	*t*	*df*	Sig. (2-tailed)	Mean difference	Std. error difference	Lower	Upper
AI-ELL activities to improve speaking skills	Equal variances assumed	0.007	0.932	−2.764	111	0.007	−0.592	0.214	−1.016	−0.168
Survey Question 15. About following features of AI tools for EAP speaking practice, indicate how important you think they are	Equal variances not assumed			−2.879	68.887	0.005	−0.592	0.206	−1.002	−0.182

Table 17.2 shows $F = 0.007 < 0.05$, whereby the assumption of equal variance and the assumption of normality were violated. However, these data still have marginal significance in virtue of Levene's test (Pallant, 2013) being at $0.932 > 0.05$, considering the first line where $p = 0.007$ (2-tailed). In this tailed independent sample *t*-test, the Sig (1-tailed) should be calculated by Sig (2-tailed) divided by 2. The $p = 0.0035$ (1-tailed) results as <0.05, which reveals how Y1 seemed to hold stronger attitudes towards AI-ELL app efficiency to improve their speaking skills than did Y2.

H2: Frequency of AI-ELL Apps' Use and Attitudes About AI-ELL Apps' Efficiency

Participants rated their attitudes from 1 to 6 on a Likert scale to express agreement on whether AI-ELL apps' use improved their English-speaking skills. The mean for the attitude test was $M = 3.01$(SD = 0.94 $N = 113$), suggesting that while 24.8% of the sample believed this statement, 6.2% of the sample did not. This suggests that students using AI-ELLs apps more frequently hold stronger beliefs in their efficiency to improve student speaking skills.

H3: Non-parametric Statistics Comparing University Attendance and Attitudes About AI Tools Related to Replacing Face-to-Face Teaching

Participants were asked to rate on a scale of 1–3 whether they thought that AI-ELL apps could replace face-to-face teaching. Number 1 represented participant belief that AI-ELL apps could replace face-to-face teaching, 2 that they could not, and 3 gave participants the opportunity to enter another opinion.

Figure 17.2 illustrates the mean for the attitude test that was $M = 1.81$(SD = 0.392 $N = 112$). While 18.6% of the sample thought that AI tools could replace face-to-face teaching, 80.5% thought they could not. The results from the Chi Square (Pallant, 2013) interestingly seemed also to suggest that a relationship between a lack of university attendance and positive attitudes that AI-ELL apps could replace face-to-face teaching existed. These results nevertheless did not represent the student

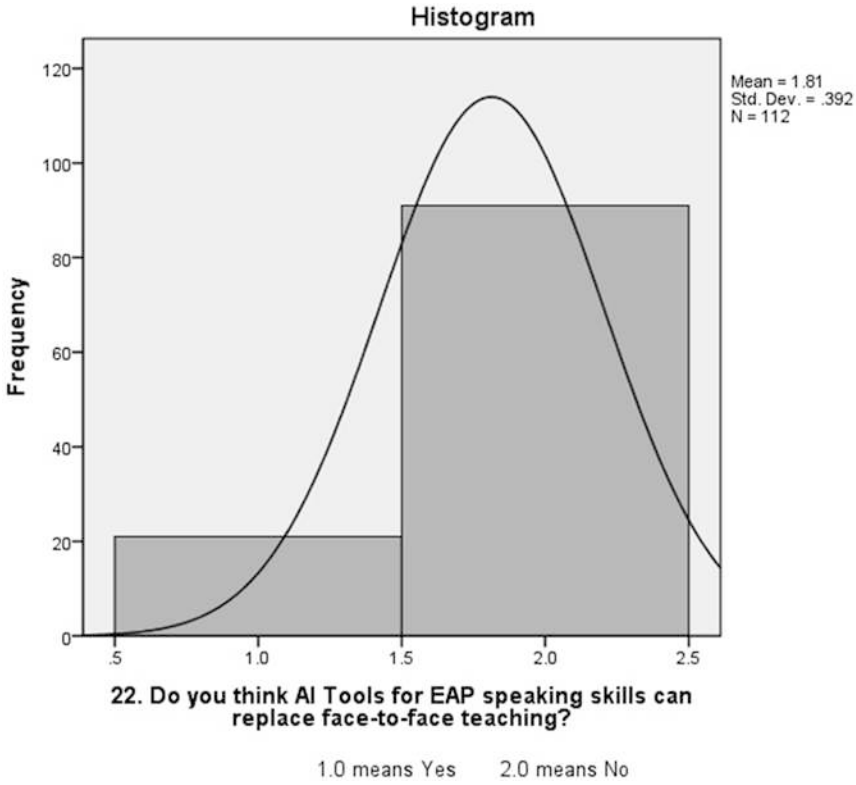

Fig. 17.2 H3: Student attitudes towards the possibility of AI-ELL apps to replace face-to-face teaching

population as a whole, which held a strong opinion that English language teachers were indispensible.

H4: Non-parametric Statistics Comparing Year of Attendance and Attitudes About AI-ELL App Costs Limiting Student Access

Participants were asked to rate from 1 to 5 the impact that price had on limiting student access to AI-ELL apps (1-definitely; 2-very probably; 3-probably; 4-possibly; 5-probably not). The mean for the attitude test was M = 2.78(SD = 1.14 N = 113). While 38.9% of the sample thought

Table 17.3 H4: Mann-Whitney between year of attendance and attitudes towards AI-ELLs price

Hypothesis test summary			
Null hypothesis	Test	Sig.	Decision
1 The distribution based on Price (if not affordable or acceptable) Survey Question 21: Do you think price is a limiting factor for AI-ELL speaking skill apps? (Likert scale applied) Please rank each possibility. Please indicate your year in university	Independent-samples Mann–Whitney *U* test	0.836	Q 21 retains the null hypothesis

Asymptotic significances are displayed. The significance level is 0.05

that price was limiting students' ability to rely on AI-ELL apps to improve their speaking skills, 7.01% of the sample thought that price was not discouraging. Moreover, the output of Mann-Whitney U test in Table 17.3 reports 0.836 > 0.05, confirming no significant difference between a student's year at the university and attitude towards price limitation to access AI-ELL apps, as observed in Chen (2011).

Qualitative Data Analysis

Participant interview excerpts are faithfully reported here without any alterations. Consequently, some inaccuracies may occur in student interview responses. Although interviews were mainly in English, Chinese translation was amply available to participants thanks to the assistance offered by project-linked students. Moreover, responses are faithfully reproduced here as copied from the coded interview transcription; the highlighted text appears where nodes were identified and coding applied (Thomas, 2006).

Students' Knowledge of Current AI-ELL Apps' Availability

One of the popular AI-ELL apps for English-speaking practice among participants was Liulishuo. Although other AI-ELLs apps for English-speaking practice were mentioned, such as Shanbay (no citation avail-

able; transcribed literally), Duolingo (2018) and Baicizhan (2018), these apps do not offer AI-supported interaction. Overall, participants reported limited knowledge of AI-ELL apps currently in the market and were unaware of any currently available apps for English for academic purposes (AI-EAP). As three students said below:

- **S1(M)**: I don't know much about the AI tools for English speaking practice.
- **S2(M)**: I just know Liulishuo, I can't think of any other product for the moment.
- **S3(F)**: Maybe know some. I cannot remember the names.

Reasons for Using AI-ELL Apps

Although one student (S5 (M)) reported that the English-speaking environment at the project university provided sufficient exposure to English, most reported not using AI-ELL apps, preferring face-to-face practice with EAP tutors. However, participants reported this was very limited in class, leading to very limited progress. Hence, most participants would prefer to practice EAP speaking skills individually by means of reading and memorizing from texts (S1 (M)). Limited progress was also reportedly achieved by practising with other students, who would not be able to identify mistakes or inappropriate use of vocabulary, or to translate from Chinese into English (Sun, Branum-Martin, Peng, & Tao, 2018).

- **S5(M)**: The main reason for not using it is because there are not many products, and finding them is troublesome. The second point is that in campus, it's more efficient to find someone to practice English face-to-face without wasting time on looking for AI products.
- **S1(M)**: I was mainly improving myself by reading, memorising some written articles or short essays … I feel objectively that the help given by the teacher is not as fast as doing something by myself.

Students' Perceived Difficulties in Making Progress with EAP Speaking Skills

Participants reported concerns about understanding the grading criteria used for EAP speaking exams. Although grading descriptors were provided, students described feeling unsure that examiners used them consistently and were consequently confused when preparing for EAP speaking exams (S2(M)). Moreover, participants felt they lacked model examples—from teachers or applications—of speech performance considered as high-scoring marks in the EAP speaking assessment.

- **S2(M)**: I think it's because the examiner, so I don't know his grading criteria, so I don't have a target when I practice this.

Students' Perceived Advantages of AI-ELL and AI-EAP Apps

The main reported advantage was AI-ELL apps' portability, followed by help with fluency and vocabulary learning—although not subject-specific—but that no single app afforded all EAP assessment-relevant skills. Participants who were familiar with AI-ELLs apps reported being satisfied with the feedback on vocabulary, grammar, speaking pace and intonation, but only in the absence of tutor feedback in class. As S2 (M) noted:

- **S2(M)**: The first thing is that if you talk to a real person, he may not always correct your pronunciation or accent, unless he's a teacher. But AI can give feedback on that. Though you may speak in a weird accent, AI can at least tell you what you pronounce wrongly. But if you talk to a real person in a face-to-face manner, you can actually improve your English skills only if the speaker's oral English level is higher than you.

An unanticipated finding was participants' concern with their inhibition to speak. Students felt too timid to speak to tutors or in public and reported that using AI-ELL apps lessened their self-consciousness, as reported below:

- **S3(F)**: But if you are really shy, you may think it's easier to talk with AI.
- **S4(M)**: Some people have psychological barriers, I think AI should be helpful.

Students' Perceived Limitations of AI-ELL and AI-EAP Apps

Many participants reported a problem with poor voice recognition, in line with evidence from Kim (2006), and Young and Mihailidis (2010). Consequently, AI-ELL apps could not provide reliable feedback on speech performance, often because despite accuracy, non-native accents were not recognized. For example, S5(M) reported:

- **S5(M)**: [I am] satisfied: the AI apps will give you more information and more detailed feedback to improve your spoken language. Dissatisfied: it may not be very intelligent, and sometimes the voice recognition is not very accurate: you said this word, but it may be recognized as another word.

Students felt that voice recognition accuracy needed particular attention as this was regarded as a decisive feature for AI-ELL/AI-EAP app efficiency. The availability of accurate, EAP-specific feedback was as desired by participants as it was perceived as not being properly addressed by AI-ELL apps. Moreover, pronunciation, accent, fluency, vocabulary, delivery and content were not observed to be comprehensively evaluated by a single app, compounded by the poor voice recognition impairing the

quality of feedback. Additionally, logical, coherent speech organization was observed as an important feature not provided by any currently available AI-ELL app.

Discussion

Starting with a general assessement of the participants, it is apparent that Y1 students rely on AI-ELL apps for speaking skills more than Y2 students do, a finding that can be explained by the newer students not wanting to feel so exposed as when speaking in public in English. Liulishuo turned out to be the most-used AI-ELL app in the Chinese market, as its features allow users to do exercises and receive feedback on vocabulary learning and fluency. Students found it difficult to conciliate their need to practise speaking skills, feeling self-conscious when speaking in a foreign language. Although preferred, access to face-to-face practice with EAP tutors was limited, as was the feedback available during classroom activities. Students also considered practising with peers as the English-speaking campus might afford more opportunities of interacting with speakers of English; however, participants reported concern that peer feedback might be deficient as a means to evaluate appropriate use of vocabulary, identify inaccurate structures or offer proper guidance to meet EAP speaking assessment requirements. Nevertheless, what the apps had to offer students was not much better. Participants reported the lack of model examples that could provide them with accurate feedback of how to successfully meet EAP exam requirements.

Students generally disagreed with the possibility that AI-ELL apps might replace face-to-face teaching, especially given their current limitations regarding their ability to recognize non-native English speakers' pronunciation properly, thus penalizing accurate performance on account of poor voice recognition. AI-ELLs apps' portability and their affordance to practice outside class were seen as potentially compensating for the limited access to tutor feedback and the limited experience peers might have in giving accurate, EAP-tuned feedback. Participants indicated AI-EAP apps would be worth the expense, provided that they could comprehend activities reinforcing grammatical accuracy, speech organization

for presentation tasks, communicative skills needed to participate in group discussions and academic vocabulary: features and skills necessary to match EAP speaking assessment tasks. Importantly, AI-EAP apps should be available offline and be able to recognize non-native English speakers' pronuniation.

Despite the limitations of time and resources, the questionnaire findings originated lines of qualitative enquiry that were further supported by the evidence found in the literature reviewed, which suggested that there is a need for more committed deployment of resources aimed at studying how AI-EAP apps can help students improve their speaking skills by matching their assessment needs, their largely non-native speaking status and financial constraints.

Conclusions

This study has explored students' attitudes towards using AI for language learning, specifically, EAP for speaking. The findings indicate that, overall, participants had positive comments on the AI technology for speaking development, although there are some limitations. The present study is to be considered a preliminary step towards further investigation into strategies for the development of AI-EAP for speaking skills. It is hoped that this research will boost interdisciplinary studies of AI technology for EAP, contributing to scholarship at local and international levels, as research into the recognition of non-native accents has gained momentum in English-speaking academia globally, featuring English accents such as Mexican, South African and Chinese (Kim et al., 2018; Liakin et al., 2015; Murphy, 2014). Furthermore, the collection of student perspectives on specific language learning skills has the potential to inform teacher training that wishes to integrate a focused approach to relevant technology for EAP teaching and learning.

It is recommended that further research maintains an interdisciplinary perspective to uncover further knowledge gaps, in order to explore and reinforce the relevance of its results in a field so far marginally explored in HE in local, Chinese and international contexts.

Limitations

The plan to run comparisons between Chinese and foreign students was impeded by sample limitations of other nationalities' pronunciation abilities. Moreover, an even distribution of the sample might have improved both the validity and reliability of the report, and consequently the impact of its findings. This will be paramount should the tool be refined in future studies. Equally, time constraints and limited resources meant that gender background data were not collected; however, this would have yielded further descriptive data on student AI-ELL apps. In the future, a clearer distinction between AI-ELL apps and AI-EAP apps in questionnaire items would yield more focused data on student preferences and perceptions. However, the study's results are relevant if considered within the limited available data on this field, backgrounding experiments that test AI-EAP apps to be released in Chinese higher education. With these considerations in mind, we believe that the present research represents a very positive step forward in understanding the benefits and limitations of AI applications aimed at language learners. It is hoped that these findings are considered a road map to AI app developers as well as a reminder that at the heart of any good language learning app is the heart of the language learners themselves.

Acknowledgement This research is supported by KSF-E-16 in XJTLU.

Glossary

Acronyms and abbreviations	Transcription and explanations
AI	Artificial Intelligence.
AI-EAP	Computer or mobile phone applications using artificial intelligence to support the learning of English for academic purposes.
AI-ELL	Computer or mobile phone applications using artificial intelligence to support the learning of English for generic purposes, as opposed to specific ones, such as academic, or aimed to prepare for high-stake exams like TOEFL iBT, and similar exams releasing certification.
CALL	Computer-assisted language learning.
EAP	English for academic purposes. University courses offered typically to non-native speaking students to ensure their familiarity with genre and text requirements of academic work designed and communicated through the medium of English.
ELF	English as a lingua franca; the dialect typically used for academic purposes by international universities.
IELTS	International English language testing system.
NNA	Non-native accents; emerging varieties of English for academic purposes such as Chinese, Mexican and Nigerian English.
NS	Native students/speakers of the target languages.
NNS	Non-native students/speakers of the target language.
TOEFL iBT	Test of English as a Foreign Language, Internet-Based Test.
TOEIC	Test of English for International Communication.
VRT	Voice recognition technology and software.

References

Al-Fahad, F. N. (2009). Students' attitudes and perceptions towards the effectiveness of Mobile learning in King Saud University, Saudi Arabia. *Turkish Online Journal of Educational Technology, 8*(2), 1–9.

Apple Inc. (2018). *Siri* [Software]. Retrieved from https://www.apple.com/siri/

Baicizhan. (2018). *Baichizhan* [Software]. Retrieved from https://www.baicizhan.com/

BALEAP. (2018). *BALEAP Can Do Framework*. Retrieved from https://www.baleap.org/resources/can-do

Beaven, A., & Neuhoff, A. (2012, January 1). *Assessing oral proficiency for intercultural professional communication: The CEFcult project*. European Association for Computer-Assisted Language Learning (EUROCALL). Retrieved from https://eric.ed.gov/?id=ED544436

Bernstein, J., Cohen, M., Murveit, H., Rtischev, D., & Weintraub, M. (1990). *Automatic evaluation and training in English pronunciation*. Paper presented at Conference. ISCA, Kobe, Japan, pp. 1185–1188. Retrieved from https://www.isca-speech.org/archive/icslp_1990/i90_1185.html

Boersma. P., & Weenink, D. (2018). *Praat: Doing Phonetics by Computer* [Software]. Retrieved from: http://www.fon.hum.uva.nl/praat/

Bordonaro, K. (2003). Perceptions of technology and manifestations of language learner autonomy. *CALL-EJ Online, 5*(1). Retrieved from http://callej.org/journal/5-1/bordonaro.html

Bruce, I. (2011). *Theory and concepts of English for academic purposes*. London: Palgrave Macmillan.

Cargill, M., & O'Connor, P. (2011). *Writing scientific research articles: Strategy and steps*. London: Wiley-Blackwell.

Celce-Murcia, M., Brinton, D. M., Goodwin, J. M., & Griner, B. (2010). *Teaching pronunciation. Hardback with audio CDs (2): A course book and reference guide*. Cambridge: Cambridge University Press.

Charmaz, K. (2003). Q*ualitative interviewing and grounded theory analysis*. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interview research: Context & method* (pp. 675–694). London: Sage.

Chavan, K., & Gawande, U. (2015). Speech recognition in noisy environment, issues and challenges: A review. In *Proceedings of 2015 International Conference on Soft-Computing and Networks Security (ICSNS)*. Coimbatore.

Chen, H. H.-J. (2011). Developing and evaluating an oral skills training website supported by automatic speech recognition technology. *ReCALL, 23*(1), 59–78.

Chiu, T.-L., Liou, H.-C., & Yeh, Y. (2007). A study of web-based oral activities enhanced by automatic speech recognition for EFL college learning. *Computer Assisted Language Learning, 20*(3), 209–233.

Chivox, Ltd. (2018a). *CHIVOX-kami English system [computer software]*. Suzhou: Chivox Ltd.

Chivox, Ltd. (2018b). *CHIVOX-Pioneer in intelligent speech analysis technology [computer software]*. Suzhou: Chivox Ltd.

Council of Europe. (2018, April 16). *Common european framework of reference for languages: Learning, teaching, assessment (CEFR)*. Retrieved from https://www.coe.int/en/web/common-european-framework-reference-languages

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: SAGE Publications.

de Jong, J., & Benigno, V. (2018, April 16). *The CEFR in higher education: Developing descriptors of academic English*. Retrieved from http://ukalta.org/wp-content/uploads/2016/10/DeJongBenigno_LTF2016.pdf

Demouy, V., & Kukulska-Hulme, A. (2010). On the spot: Using mobile devices for listening and speaking practice on a French language programme. *Open Learning: The Journal of Open, Distance and e-Learning, 25*(3), 217–232.

Deng, Q., & Trainin, G. (2015). Learning vocabulary with apps: From theory to practice. *The Nebraska Educator, 2*, 49–69.

Derwing, T. M., Munro, M. J., & Carbonaro, M. (2012). Does popular speech recognition software work with ESL speech? *TESOL Quarterly, 34*(3), 592–603.

Dlaska, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System, 36*(4), 506–516.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford Applied Linguistics.

Douma, P., Anderson, G., Akahane, M., & Mizikovsky, S. (1996). Methods and apparatus for training and operating voice recognition systems. In *US5583965A Documentation*. Retrieved from https://patents.google.com/patent/US5583965A/en

Duolingo. (2018). *Duolingo: Learn Spanish, French and other languages for free [computer software]*. Cheshire: Duolinguo.

Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.

Gilakjani, A. P. (2011). A study on the situation of pronunciation instruction in ESL/EFL classrooms. *Journal of Studies in Education, 1*(1:E4), 1–15.

Gilakjani, A. P., & Sabouri, N. B. (2016). How can EFL teachers help EFL learners improve their English pronunciation? *Journal of Language Teaching and Research, 7*(5), 967–972.

Glasman-Deal, H. (2010). *Science research writing for non-native speakers of English*. London: Imperial College Press.

Google Cloud. (2018). *Cloud Speech-to-Text API* [computer software]. Retrieved from https://cloud.google.com/speech-to-text/

Hincks, R. (2005). Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism. *System, 33*(4), 575–591.

IBM SPSS. (2018). *IBM SPSS. Statistics package for the social sciences* (Version 22) [Software]. Retrieved from http://www.ibm.com/analytics/us/en/technology/spss/

IELTS. (2018). IELTS. Australia: British Council. Retrieved from https://www.ielts.org/

iFlytek Co., Ltd. (2018). *iFlytek* [Software]. Retrieved from http://www.iflytek.com/en/index.html

Jenkins, J. (2014). *English as a lingua Franca in the international university: The politics of academic English language policy*. Oxon: Routledge.

Jenkins, J. (2017). *Mobility and English language policies and practices in higher education*. Oxon: Routledge.

Jia, J. (2009). An AI framework to each English as a foreign language: CSIEC. *AI Magazine, 30*(2), 59–71.

K12 Inc. (2018). *K12 online education programs & schooling*. Retrieved from https://www.k12.com/

Kan, Q., & Tang, J. L. (2018). Researching mobile-assisted English language learning among adult distance learners in China: Emerging practices and learner perception of teacher role. *International Journal of Computer-Assisted Language Learning and Teaching, 8*(3), 1–28.

Kang, O., Thomson, R., & Moran, M. (2018). The effects of international accents and shared first language on listening comprehension tests. *TESOL Quarterly, 53*(1), 56–81.

Kessler, G., Bikowski, D., & Boggs, J. (2012). Collaborative writing among second language learners in academic web-based projects. *Language Learning and Technology, 16*(1), 91–109.

Kim, I.-S. (2006). Automatic speech recognition: Reliability and pedagogical implications for teaching pronunciation. *Journal of Educational Technology & Society, 9*(1), 322–334.

Kim, Y., Soyata, T., & Behnagh, R. F. (2018). Towards emotionally aware AI smart classroom: Current issues and directions for engineering and education. *IEEE Access, 6*, 5308–5331.

Knight, W. (2017). *China's AI awakening. The West should stop worrying about China's AI revolution*. Retrieved from https://www.technologyreview.com/s/609038/chinas-ai-awakening/

Köse, U., & Arslan, A. (2014). Design and development of a chaos-based image encryption system. In S. Banerjee & Ş. Ş. Erçetin (Eds.), *Chaos, complexity and leadership 2012* (pp. 23–28). Dordrecht: Springer.

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal, 32*(1), 1–25.

Liulishuo. (2017). *Liulishuo—Your personal AI English teacher* [Software]. Retrieved from https://www.liulishuo.com/en/aboutus.html

Mauranen, A. (2012). *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge: Cambridge University Press.

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System, 57*, 25–42.

Meisam, R., & Tavakoli, M. (2015). The effectiveness of CALL in helping Persian L2 learners produce the English vowel /ɒ/. *GEMA Online Journal of Language Studies, 15*(3), 17–30.

Murphy, J. M. (2014). Intelligible, comprehensible, non-native models in ESL/EFL pronunciation teaching. *System, 42*, 258–269.

Neri, A., Cucchiarini, C., & Strik, W. (2003). *Automatic speech recognition for second language learning: How and why it actually works*. Paper presented at 15th ICPhS Barcelona, Spain.

Nuance Communications. (2018). *Dragon naturally speaking* [Software]. Available from https://www.nuance.com/dragon.html

Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement*. London: Continuum.

Pallant, J. (2013). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Maidenhead, Berkshire: McGraw-Hill Education.

Park, M., & Slater, T. (2014). A typology of tasks for mobile-assisted language learning: Recommendations from a small-scale needs analysis. *TESL Canada Journal, 31*(SI8), 93–115.

Reinders, H., & Darasawang, P. (2012). Diversity in learner support. In G. Stockwell (Ed.), *Computer-assisted language learning: Diversity in research and practice* (pp. 49–70). Cambridge: Cambridge University Press.

Sapsford, R. (1999). *Survey research*. London: SAGE Publications.

Setter, J., & Jenkins, J. (2005). State-of-the-art review article. *Language Teaching, 38*(1), 1–17.

Sun, C., Branum-Martin, L., Peng, P., & Tao, S. (2018). Phonology, orthography, and decoding skills within and across English and Chinese. *Scientific Studies of Reading, 22*(5), 401–419.

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237–246.

Wang, Y.-H., & Young, S. S.-C. (2014). A study of the design and implementation of the ASR-based iCASL system with corrective feedback to facilitate English learning. *Journal of Educational Technology & Society, 17*(2), 219–233.

Xu, Q., & Peng, H. (2017). Investigating mobile-assisted oral feedback in teaching Chinese as a second language. *Computer Assisted Language Learning, 30*(3–4), 173–182.

Young, V., & Mihailidis, A. (2010). Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review. *Assistive Technology, 22*(2), 99–112.

Zhang, H., Song, W., & Huang, R. (2014). Business English vocabulary learning with mobile phone: A Chinese students' perspective. *International Journal of Computer-Assisted Language Learning and Teaching, 4*(2), 46–63.

Zou, B., Li, H., & Li, J. (2018). Exploring a curriculum app and a social communication app for EFL learning. *Computer Assisted Language Learning., 31*(7), 694–713.

Zou, B., Wang, D. S., & Xing, M. J. (2016). Collaborative tasks in wiki-based environment in EFL learning. *Computer Assisted Language Learning, 29*(5), 1000–1016.