Chinese University Students' Cognition of Mobile Speech Recognition Software in English Communicative Training

Xiaomeng Li

Liyxy171@mymail.unisa.edu.au

Master of Education (TESOL), UniSA Education Futures

Abstract. With the popularity of intelligent artificial intelligence (AI)-based education programs, many Chinese education scholars have found opportunities in communicative English learning. However, there is little qualitative research on the use of automatic speech recognition technology in English language-learning applications. Therefore, based on a week of user experience with a mobile app for self-study English called Liulishuo, the author designed a case study to explore how Chinese university students perceive the app. The author used the example of Chinese students who used Liulishuo and asked how it can improve pronunciation skills and speaking ability. The author collected a variety of data from participants' biographical discourses, learning journals, and personal interviews. The results of the thematic analysis using Dedoose qualitative research data analysis software showed that most of the participants enjoyed learning experiencesin using Liulishuo and would continue adopting it in future self-study.

Keywords: Automatic speech recognition, case study, communicative English, universitystudents, mobile-assisted language learning

1   Introduction

In 2017, the average English proficiency score of Chinese English learners in Shanxi province was 50.67, ranking second to last in China, while Shanghai ranked first with 57.91 [1]. As stated by Ahn and Lee [2], for most English as a foreign language (EFL) learners, the most challenging English skill is speaking. Communicative English learning, especially localized English courses in Chinese university education, is taught in classrooms featuring grammar translation and teacher-centered instruction. As recipients of knowledge, Chinese learners are rarely exposed to learning environments where English is the language of instruction, and they rarely have the opportunity to speak in the target language [3]. Pronunciation exercises suffer from similar problems, although pronunciation "permeates all spheres of human life, lies at the core of spoken expression and embodies the ways in which speaker and listener work together to co-produce and understand each other" [4].

1.1 Benefits of mobile phones in communicative English learning

The portable, easy-to-use, and practical features of mobile devices are conducive to the development trend of mobile-assisted language learning (MALL). In February 2021, the statistics on China's mobile phone ownership showed that about 1.6 billion Chinese subscribed

to cell phone numbers [5]. This means that Chinese learners generally use smartphones and related apps [2]. Therefore, some researchers have asserted that apps that support communicative English learning can motivate those who are shy in public to actively participate in different social and cultural contexts [2]. Ultimately, portable phones may integrate learning content into contextual learning, both inside and outside the smart classrooms [1] [6] [7]. In fact,many scholars have emphasized how speech-activated apps built with automatic speech recognition (ASR) can be used to maximize oral learning performance [7] [8] [9] [10]. As a result, education researchers have witnessed a range of English language learning (ELL) software that uses ASR to facilitate language learning.

Apps using ASR are conducive to enriching socialization in virtual learning environments, and they feature authentic self-learning resources. A comfortable learning space without cognitive overload can create an interactive and collaborative community with native speaker peers, reducing learners' fear of speaking in front of others [3]. In addition, Rao [3] suggested using a self-paced learning approach to engage students in various speaking activities in chronological order of learning, practicing, and assessment.

1.2 Case study, research objective, and research questions

There are few studies on the potential of MALL for ASR applications. This case study aims to bridge the research gap regarding Chinese university students as users of ASR-equipped ELL speaking apps. Mezirow's transformative learning theory provides the conceptual framework. The author conducted a case study of Chinese undergraduate EFL students with the research goal of reflecting students' critical self-awareness in terms of satisfaction with their experiences.

With the permission of a university gatekeeper in Shanxi province, the author utilized data from 'triangulated' sources: participants' biographical utterances, a week's study log, and personal interviews. The author started translating and transcribing data using the XunFei Tingjian computer software. Finally, thematic analysis, in Dedoose software, was used to generate manageable concepts and codes for synthesizing findings with relevant literature for discussion. There are two main research questions (RQs) and three sub-questions:

How do university-level Chinese EFL learners perceive the use of ASR-based mobile technology in acquiring oral English skills?

1.1 What can they learn from experiential pedagogy in digital classrooms?

How to improve ASR-based mobile apps for better oral English tutoring?

2.1 To what extent learners will continue to use these apps?

2.2 What future directions do they suggest?

1.3 Theoretical framework: Transformative learning

All the rationale for RQs regarding participants' self-reflection on gained knowledge and apps' potential can be derived from Mezirow's constructivist theory of transformative learning. The theory focuses on learners' *habits of mind*, mindsets, meaning perspectives, and *frames of reference*' [11]. As Mezirow [12] pointed out, there are two types of learning. A teacher's *instrumental learning* may include learning a new method to enrich learners' perceptions [11]

and helps learners' mutual instrumentalization, so that their metacognitive reasoning mindsets can be promoted.

2 Data sampling, methods, materials, and thematic data analysis

2.1 Case sampling strategy

The research was conducted at the Xi'an University of Architecture and Technology (XAUAT) in Xi'an, China. The author gained access to the site and participants by contacting the college and initiating negotiations. Obtaining the gatekeepers' endorsement and the participants' informed consent facilitated data collection [13]. First-hand data sampling uses a purposeful strategy to select key cases [14]. For this study, the focus was on students in higher education who had difficulties learning to speak English, especially pronunciation.

2.2 Triangulation and data-collection techniques

It was decided to use interviews, participants' study diaries, and biographical utterances as data sources. The biggest advantage of triangulation between techniques like this is that multiple sources of evidence help develop data corroboration and aggregate lines of inquiry [15]. Therefore, any findings or conclusions drawn using triangulation will be more convincing because it forms structural validity when describing the same social event from multiple datasets [15].

2.3 Interview guide

Following the interview guide, the author scheduled time with the participants and then recorded data via Zoom. After one-week trial use of Liulishuo, in addition to negotiating with the interviewees to schedule time, the instruction manual was also distributed to each participant in advance.

2.4 Thematic data analysis

This study used two features of qualitative analysis aimed at investigating patterns in data and generating codes that explain the phenomenon. Based on accumulating homogeneous and case-specific inductive analysis, it was allowed to generate and justify generalizable explanations [16]. The author adopted a bottom-up approach to construct patterns, categories, and themes into identified codes and frequencies to explore the university students' subjective perspectives on ASR–ELL apps. Meanwhile, the deductive analysis allowed the construction of general statements or theories as starting points for explaining specific circumstances [16]. Therefore, a top-down approach was adopted, responding to the possible influencers of the two main RQs (see Table 1 in Section 3) and the theoretical framework against recurring synthesis themes. This process was achieved by linking the data to previous relevant literature and the principles of the transformative learning theory.

# 3 Response to research questions through data collection and findings

To investigate the participants' cognition of ASR technology as it might help improve their English speaking skills, after one week's trial of Liulishuo, all five participants' biographical utterances, and personal interviews. Table 1 shows the three corroborative data sources which respond to the three sub-RQs. The response was consistent with the deductive analysis that explained similar findings from previous studies on SR-informed learning journeys.

The content of the interview was transcribed using XunFei TingJian software and analyzed using thematic analysis by the Dedoose qualitative research software. Participants' responses were coded using only the initials of their last names, thereby protecting the participants' anonymity. The thematic map of codes used to analyze interviews and journals is shown in Figure 1.

Table 1. Overview of sub-categorized RQ identified in data sources

Sub-RQ	Data sources		
	Individual interviews	Learning journals	Biographical utterances
Participants learned (or not) from the experientialpedagogy in Liulishuo	x	x	
Views on extendingtrial use (i.e., unlockingpaid version)	x	x	x
Suggestions concerning Liulishuo and similar ASR-informed speaking learning apps	x		

(Note: x means corresponding convergence in evidence to address sub-RQ)
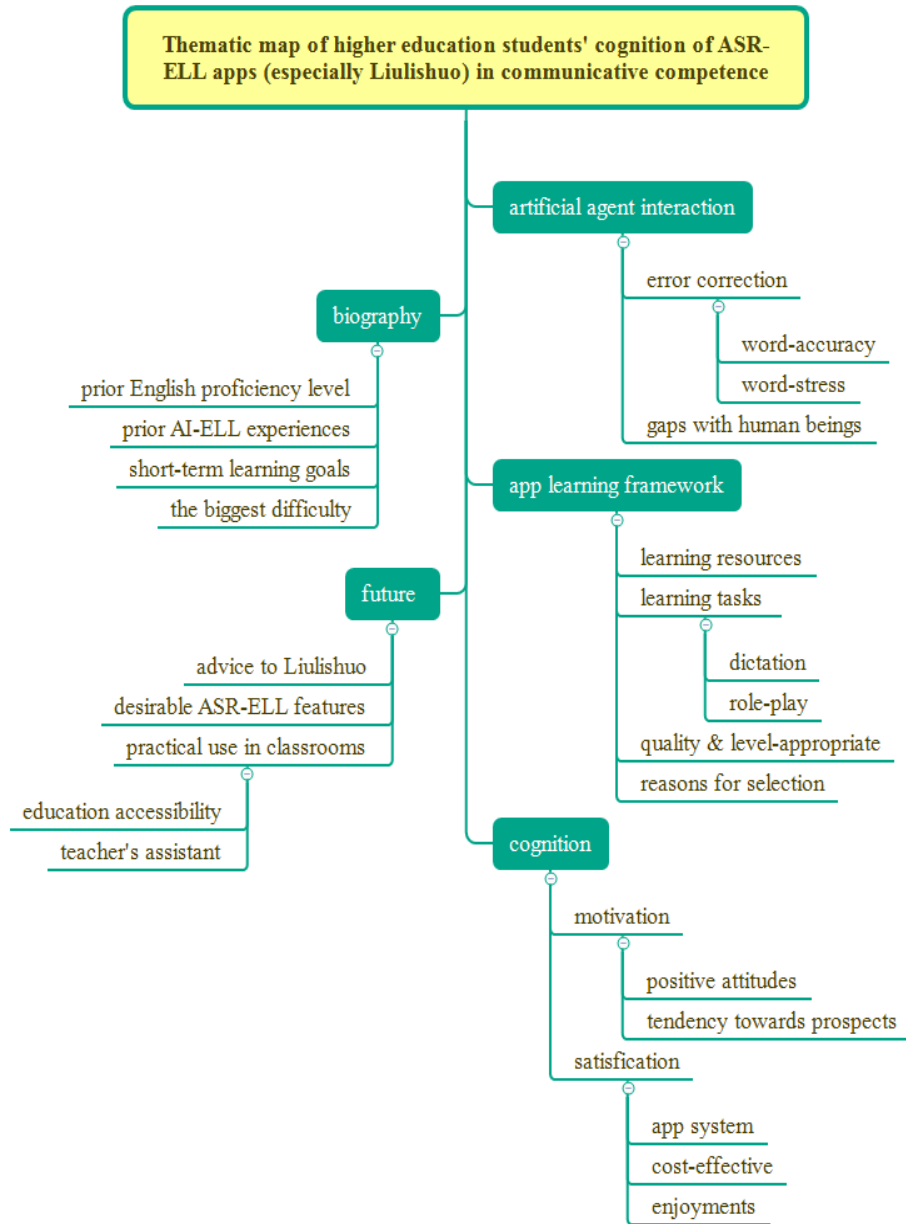
Fig 1. Thematic map of participants' cognition of ASR-ELL apps in communicative competence

3.1 Biography

Five male participants were majoring in engineering. They were honors students at undergraduate level at XAUAT. Table 2 shows the biographically relevant themes from the interviews.

Table 2. Participants' biographical data

Participant (named by initial letter of last names)	Academic background	Age	English proficiency level	Technology-assisted English learning experiences (recent months/past year)	Preference in learning settings (virtual vs. in-person)	Short-term goal for speaking skill	Biggest challenge with speaking English
L	Civil engineering	20	IELTS 5.5 (speaking)	Baici zhan English vocabulary	In-person	IELTS 7	Fluency
H	Software engineering	21	CET-4	Shanbei English vocabulary	In-person	No	Transfer Chinese mindset into English mindset
S	Civil engineering	20	CET-6	Liulishuo & Tandem	In-person	No	Fluent speech
G	Civil engineering	21	IELTS 6.0 (overall and speaking)	Kaiyan English speaking	In-person	Understand native speakers' dialogues	Slang & idiomatic usages
Y	Electronic engineering	21	IELTS 6.5 (overall)	No	Either	Communicate freely with native speaking business partners in English (3 months)	Inaccurate pronunciation

3.2 Learners gained knowledge and skills in ASR–ELL apps

Liulishuo captured three highlights of the learning moments, and the most frequent themes are 1) 'good-quality, level-appropriate' learning content, reflected in 2) 'learning tasks' and 3) rich 'learning resources.' In placement tests, most of the participants stated that their scores were likely to estimate speaking proficiency with the interview sections, such as 'no overlong sentences; easier for me to follow' (H), and 'speech recognition analyses and identifies common problems' (Y). The dictation tasks were chosen as the most useful learning exercises in an enjoyable learning experience. For example, L wrote in his learning diary that the most frequently practiced oral task in 10 to 15 minutes every day was 'Chinese to English oral translation.' Likewise, H also spent 10 minutes reviewing the monologue because 'I stick to syllables or words that I feel difficult to pronounce and will repeat several times, and then I can speak shadowing segments.'

In addition to the task-based learning built into the Liulishuo system, the app also launched its complementary app named Liulishuo Pie, which provides fruitful learning resources to immerse learners in a real learning environment; the value of these resources was confirmed

by all participants in the interviews. For instance, there were audio-visual learning materials extracted from 'films, television clips, and songs in English' (S).

3.3 Continuing use after the trial

Themes in addressing the second sub-RQ on consistent usage of Liulishuo were interpreted in the interview data and coded as 'tendency towards prospect.' Four of five participants desired to subscribe as paid membership of Liulishuo. This was reflected in short remarks such as 'I am going to unlock further learning units' (Y; S). The supplementary app Liulishuo Pie, which features role-playing learning tasks to communicate with English speakers from foreign countries, attracted Y and S's huge interest in becoming paying users. In the journal data of participants L and G, in particular, L said in an interview he would not subscribe to Liulishuo and that they might stop using such apps. For example, both L and G complained that they performed poorly on the Chinese-English translation task at level 2 (which corresponded to IELTS 5.5–6.0) because their scores would be higher only if they filled in the exact word choice that matched with the unique answer. Therefore, it means that they seemed unlikely to use Liulishuo for further communicative learning.

3.4 Suggestions for ASR–ELL apps

Responses to the top-level theme 'future' answered the third sub-RQ about future developments of Liulishuo and ASR speaking and learning apps. Interviewees Y, H, and S suggested there should be more interest-oriented, contextualized speaking topics (Y; H) as well as spoken vocabulary exercises (Y), improved speech recognition technology (H), and less frequent mid-term unit tests (S). Their comments suggested that the designers of Liulishuo and other digital pedagogy tutors might consider users' voices and provide better service toward achieving their learning goals. Tailored learning received general comments about what apps developers should be aware of.

All participants demonstrated the practical value of these apps by adopting them in the classroom. Liulishuo and Liulishuo Pie represented expansions of accessible learning materials in the form of assistants of teachers. Directly indicative data were collected for the most frequently mentioned theme, 'education accessibility' when interviewees collectively suggested that these apps 'could increase opportunities to communicate with native English speakers virtually' (Y; G; S). Two interviewees considered Liulishuo as 'widening horizons through immersion in an English learning environment with lots of learning resources like world news, songs, and so forth' (H; Y). This piece of evidence shows that Liulishuo or other ASR-ELL apps with the same learning features can enrich Chinese university-level English learners, especially those living in rural areas of China. Additionally, interviewees G and Y felt that the ASR apps could support face-to-face English-speaking classes. In particular, G attended many in-person oral courses at the New Oriental English Tutoring Center. Because of the high cost of one-to-one tutoring, such apps can serve as self-study resources, with regular remedial assignments that adapt to learners' studying paces, he said.

4 Discussion

4.1 Return to the conceptual framework

Mezirow [11] claims that when students negotiate new and novel learning experiences through frequent engagement, they can begin to expand their frames of reference (e.g., beliefs, expectations, and mindsets) and change their perceptions, either in-depth or in relation to preconceived viewpoints. According to the interviews and biographical details, this learning goal was achieved through continuous adoption of Liulishuo based on their satisfaction experience.

4.2 Acquired knowledge and skills in experiential pedagogy

In the first main RQ (including the first sub-RQ), participants' satisfaction with the ASR notification apps for learning oral skills is reflected in Liulishuo's effective error correction, authentic learning content, and useful dictation activities. On the first-level coded 'AI interaction' theme, all participants agreed that ASR apps were beneficial for correcting mistakes in monologues and automatically recommending steps for improvement. For example, both L and Y praised the Liulishuo agent for highlighting phonetic errors as well as better vocabulary choices in the 'role-play' learning exercises using different colors on the feedback panel. After pointing out such issues, it suggested the next steps and learning objectives. These points are consistent with the literature on the effectiveness of intelligent agents in finding learners' pronunciation and vocabulary errors [17].

Second, ASR-based techniques provide multiple options for interest-oriented learning content [17]. In particular, Liulishuo provides a large number of digital learning materials that are not limited by time and space [7]. For instance, G said that because he likes to watch English speakers' talk shows, Liulishuo bridged him to the outside world on his favorite learning topics, and the price is affordable for use in his spare time in the dormitory.

Finally, the usefulness of the 'listen and repeat' learning activity in ASR–ELL apps was valued for students' pronunciation skills [17]. This was synthesized from the comments of participants L, H, and Y, who described the dictation task as featuring Mandarin-English paragraph translation. Good learning performance would be measured on a high-score evaluation mechanism after several repetitions under phonological self-correction. Therefore, given that most of the participants had low-to-intermediate past English proficiency (see Table 2), provision for the use of Mandarin as the first language for the AI agent may reinforce basic pronunciation knowledge.

4.3 Continued adoption of ASR-ELL apps

On the second sub-RQ of this study, since many participants showed positive attitudes towards Liulishuo, it can be inferred that they will continue to use it in the future, which is consistent with what is said in the literature that ASR technologies are helpful for pronunciation practice [8]. More importantly, participants' desire for continued app use was reflected in the integration with classroom teaching (in the theme' future practical use – teacher's assistant'). Continuing use of Liulishuo appeared in direct quotes from respondents about unlocking the paid version and subscribing to Liulishuo Pie (the promoted simulation ELL app). Sample phrases included 'definitely purchase it after the trial' (H and Y); 'absolutely pay for it if it's

useful in one-to-one private English-speaking education' (S and G); and 'buy Liulishuo PIE because of its peer interaction with native-like English speakers' (L). These examples imply that some participants were strongly willing to adopt the apps after trial uses. Similar conclusions can be inferred from biographical data. For example, both Y and G demonstrated their short-term learning goals of communicating in English with NS by listening to authentic English dialogues multiple times. Therefore, they will continue to use Liulishuo. Meanwhile, limited tutor access might hinder the continuity of learners' use of the apps [18]. This underscores the pedagogical value of having a freestyle conversation with peers or teachers for more appropriate performance evaluations [7].

4.4 Advice for the development of ASR-ELL apps: Voice recognition engine, UI, and feedback

After coding and investigating the themes of 'advice for Liulishuo' and 'desired features of ASR–ELL apps', it is found that answers to the third sub-RQ amounted to three pieces of advice that the apps' developers and designers should considered. They seek a higher standard of the ASR engines and more accurate feedback and correction of pronunciation issues. The first technical suggestion is to address learners' dissatisfaction with the inefficient voice-recognition features of ASR apps. All participants responded that a major hurdle preventing Liulishuo from expanding user subscriptions was the app's low error-accuracy rate. This concern is consistent with the recommendation that a less demanding ASR engine is needed to account for speech variation in non-native English speakers [8].

Another suggestion results from the fact that Liulishuo cannot provide sufficient evaluation reports. The multiple feedback issues of ASR technology may only be resolved as technology advances in evaluating learners' current progress and incrementally improving to fulfill their needs [4]. However, in addition to the efforts in digital pedagogy, learners need to develop their own self-learning skills and enable self-directed learning. Mobile learning outcomes will improve if learners are able to fully identify and assess segmental and suprasegmental errors in their speech [4]. Joint training is also useful for learners who are practicing pronunciation - for example, in terms of phonemes, rhythm, and intonation. As a result, more time can be spent on independent learning, using appropriate levels of knowledge to inform achievable learning outcomes.

5 Conclusion

This case study investigates how Chinese university students are represented by five cases of male undergraduate students in engineering considering ASR-based apps to learn oral English. Liulishuo was the unit of analysis for improving pronunciation and speaking skills. The results indicated a positive experience of using Liulishuo and continued adoption of its use in future self-learning. The reason for the enjoyable emotional experiences lies in Liulishuo's error-correction mechanisms. Liulishuo and similar ASR-ELL apps provide English-speaking improvements as assistants to classroom teachers and reveal discrepancies in education equity between rural areas and industrialized provinces. The finding of the study lack generalizability and applicability in a wider context due to the limitations of gender and age parameters. There is room for ASR research atthe pre-college levels and among female undergraduates.

Based on the technical features of the acoustic model in Liulishuo and similar ASR-ELL apps, it is recommended that developers collect and analyze as much training data as possible to improve the accuracy of the classifier and ultimately achieve reliable speech score estimation for learners [19]. It is hoped that ASR-based apps can employ speech enhancement features. As highlighted by the Chinese Academy of Science, one future direction of AI-based English learning apps, corresponding to the cases explored in this study, is pinned on personalized learning platforms [6]. This is where intelligent adaptive learning can begin to integrate cognitive science, natural language processing, deep learning, and symbolic reasoning to stimulate human-machine interaction.

References

[1] Statista.: English language proficiency in greater China in 2017, by region (in EF English Proficiency Index average scores), Statista, [ONLINE]
<https://www.statista.com/statistics/949264/greater-china-english-language-proficiency-by-   region/> (2017)

[2] T.Y. Ahn, and S.M. Lee.: 'User experience of a mobile speaking application with automatic speech recognition for EFL learning', British Journal of Educational Technology, vol. 47, no. 4, pp. 778-786,July (2016)

[3] X. Rao.: 'University Academic English: A Rising Global Phenomenon', in University English for Academic Purposes in China, Springer, Singapore, pp. 11-20 (2018)

[4] N. Bogach, E. Boitsova, S. Chernonog, A. Lamtev, M. Lesnichay, I. Lezhenin, A. Novopashenny, R. Svechnikov, D. Tsikach, K. Vasiliev, and E. Pyshkin.: 'Speech processing for language learning: a practical approach to computer-assisted pronunciation teaching', Electronics, vol. 10, no. 3, pp. 235-257, February (2021)

[5] Statista.: Number of mobile cell phone subscriptions in China from February 2020 to February 2021, Statista, viewed 17 August 2021, < https://www.statista.com/statistics/278204/china-mobile-users-by-month/> (2021)

[6] Deloitte.: 'Conversational AI is reshaping the human-machine interaction', Deloitte, viewed 28 July 2021, <https://www2.deloitte.com/cn/en/pages/innovation/articles/innovation-conversational- ai-is-reshaping-the-human-machine-interaction.html/> (2020)

[7] B. Zou, S. Liviero, M. Hao, and C. Wei.: 'Artificial intelligence technology for EAP speaking skills: student perceptions of opportunities and challenges', in Technology and the Psychology of Second Language Learners and Users, Palgrave Macmillan, Cham, pp. 433-463, (2020)

[8] Chen, HHJ.: 'Developing and evaluating an oral skills training website supported by automatic speech recognition technology', ReCALL, vol. 23, no. 1, pp. 59-78, (2011)

[9] Romanov1 N, Cluci MI, Anastasiei IC & Țugui A.: 'Artificial intelligence applications and tools in higher education: an overview', in Monnet J (ed), SESYR Sustainable Education through European Studies for Young Researchers, Alexandru Ioan Cuza University of Lasi, Romania, pp. 11-29, (2020)

[10] Wang, YH & Young, SSC.: 'Exploring young and adult learners' perceptions of corrective feedback in ASR‐based CALL system', British journal of educational technology, vol. 43, no. 3, pp. E77-E80, (2012)

[11] Bush, SB, Cook, KL, Edelen, D & Cox, R .: 'Elementary students' steam perceptions: extending frames of reference through transformative learning experiences', The Elementary School Journal, vol. 120, no. 4, pp. 692–714, (2020)

[12] Mezirow, J.: 'Transformative learning theory', in Contemporary Theories of Learning: Learning Theorists in Their Own Words, (2nd edn), Illeris K (ed), Taylor and Francis, London, pp. 114-128, (2018)

[13] Marshall, C & Rossman, GB,: Designing qualitative research, (5th edn), SAGE publications, Los Angeles, (2011)

[14] Cohen, L, Manion, L & Morrison, L.: Research methods in education, (8th edn), edited by Louis Cohen, Lawrence Manion and Keith Morrison, Routledge, London, (2018)

[15] Bashori, M, van Hout, R, Strik, H & Cucchiarini, C.: 'Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment', System, vol. 99, pp. 1-16, (2021)

[16] Gibbs, G.: Analysing qualitative data, SAGE, Los Angeles, California; London, (2007)

[17] Creswell, J.W. & Creswell, J.D.: Research design: qualitative, quantitative, and mixed methods approches, Sage Publications, Thousand Oaks, California, London, United Kingdom (2018)

[18] Waring, M.: 'Finding your theoretical position', in Arthur, J, Waring, M & Coe, R, Research Methods and Methodologies in Education, SAGE Publications, London, pp. 15-19, (2012)

[19] Marshall, C & Rossman, GB.: Designing qualitative research, (5th edn), SAGE publications, Los Angeles, (2011)