

The Impact of Instructional Videos Supported by AI-Driven Tutoring System on EFL Listening and Speaking

Xiangyu TAN^{ab*} & Xiuyuan ZUO^c

^a*School of Humanities, Shanghai Open University, China*

^b*School of English Studies, Shanghai International Studies University, China*

^c*School of Foreign Languages, University of Shanghai for Science and Technology, China*

*tanxy@sou.edu.cn

Abstract: Recent studies highlight the beneficial effects of instructional videos (IVs) and intelligent tutoring systems (ITS) on the oral development and learning attitudes of learners of English as a Foreign Language (EFL). Despite these advances, studies investigating the combined impact of AI-driven tutoring systems with instructional videos on EFL learners' listening and speaking skills remain sparse. This study aims to fill this research gap by employing a 15-week quasi-experimental design with 43 adult participants, divided into two groups. The experimental group (Group IVIT) utilized instructional videos integrated with an AI-driven tutoring system (iTutor), whereas the control group (Group IV) engaged with instructional videos alone. Comparative analysis demonstrated significant enhancements in listening and speaking proficiencies within Group IVIT, which notably outperformed Group IV in speaking skill development. Additionally, learners in Group IVIT reported more positive perceptions of the opportunities for speaking provided by IVIT and experienced less anxiety related to speaking tasks. These findings emphasize the significant potential of integrating intelligent tutoring systems with traditional learning modalities to accelerate language acquisition in adult EFL learners.

Keywords: Instructional videos, AI-driven tutor, EFL, listening skills, speaking skills

1. Introduction

Instructional videos (IVs) are multimedia resources designed specifically for educational purposes, which have become a cornerstone of online education. IVs offer dynamic, visual, and auditory learning that can simplify complex topics. Previous research highlight that IVs can providing a rich, multimodal experience that supports the acquisition of language skills (Cinganotto & Cuccurullo, 2015; Vungthong, Djonov & Torr, 2017). Research also suggests that interactive features and intelligent support are needed to ensure the effectiveness of IVs (Mitrovic et al., 2017). According to Xiao and Bai (2022), AI-driven tutoring system could

effectively support learning by adapting to individual learner needs, thus enhancing student engagement and making online learning through IVs more interactive and adaptive.

However, existing studies on integrating AI-driven tutoring system into IVs are still exploratory. Further research is needed to fully understand their potential and limitations (Leiker et al., 2023). Therefore, the current study aims to assess the effectiveness of instructional videos integrated with an AI-driven tutoring system named iTutor (IVIT) on the development of listening and speaking skills in adult EFL learners. The findings of this study could provide valuable insights into the potential of AI-driven tutors for online language learning and contribute to the advancement of our understanding in this domain.

2. Literature Review

2.1 Instructional Videos in EFL Listening and Speaking

Numerous studies have shown that using IVs can have a positive impact on the listening and speaking skills of EFL learners. Kam, Liu and Tseng (2020) found that IVs were effective in helping L2 learners comprehend and understand spoken language, thus improving their listening skills. Another study by Yükselir and Kömür (2017) highlighted the positive effects of using online videos to enhance speaking abilities in EFL learners. Similarly, Hadeel Saed et al. (2021) reported that university students found online videos to be very helpful in improving their speaking skills as they provided a more diverse and engaging language exposure compared to traditional classroom learning methods. Sun and Fang (2015) also discovered that incorporating online videos into language learning significantly boosted students' confidence in speaking English and encouraged them to practice more actively outside the classroom.

Despite the overall positive effects, concerns were also raised about passive learning problems if IVs were not properly integrated with active learning tasks that promote critical thinking and interaction (Kusumarasdyati, 2015). Therefore, intelligent supports are needed to enhance the educational effectiveness of IVs (Mitrovic et al., 2017). The use of an intelligent tutoring system (ITS) could be an effective solution to tackle these concerns.

2.2 Intelligent Tutoring System in EFL Listening and Speaking

2.2.1 Learning outcome

Previous research has shown that the use of ITS can lead to improved language learning outcomes. A meta-analysis showed that ITS-mediated instruction generally produced greater achievement compared to traditional instruction methods, supporting the effectiveness of ITS in providing personalized feedback and improving learning outcomes (Ma et al., 2014). Research has also indicated that automated, data-driven personalized feedback in ITS can significantly improve student performance outcomes (Kochmar et al., 2021).

Despite the benefits, challenges like the need for accurate modeling of complex learner

behaviors and technical issues limit the widespread adoption of ITS (VanLehn, 2011). Furthermore, the long-term impacts of ITS and its integration into standard curricula remain areas needing more robust research (Yüce, Abubakar & Ilkan, 2019).

2.2.2 Perceptions

EFL learners have also reported positive perceptions on ITS. Zou et al. (2023) found that AI-driven speech evaluation programs were promising tools to improve EFL learners' conversation and pronunciation skills. Similarly, Han's (2020) study showed that voice-based AI chatbots enhanced students' speaking abilities and offered them a comfortable learning experience. In addition, Shishido's (2021) study indicated that ITS boosted students' learning motivation, facilitated their English knowledge and skill development, and reduced their learning anxiety.

The aforementioned studies suggest that the use of IVIT might help overcome the limitations of traditional IVs. IVs that are integrated with active learning tasks such as initiation, analysis, feedback, or assessment and application of the learned material might be more pedagogically effective. However, such empirical studies are still scarce. Therefore, the current study was conducted to empirically investigate the impacts of IVIT on EFL learners' language development in listening comprehension and speaking proficiency, and the learners' perceptions about the use of IVIT. The following research questions were addressed:

- (1) Does IVIT significantly improve adult EFL learners' listening comprehension and speaking proficiency compared to IV?
- (2) How do adult EFL learners perceive the use of IV and IVIT?

3. Method

3.1 Participants

The quasi-experiment involved 43 participants who were divided into two groups: the IV group (4 male, 17 female) and the IVIT group (3 male, 19 female). These participants were enrolled in a Business English course at an open university. Their age ranged from 28 to 49 years old, with an average age of 35.72. All participants were adult EFL learners with similar proficiency levels (CEFR B1), as evidenced by their scores on mock Business English Certificate (BEC) Preliminary exams.

3.2 Instruments

3.2.1 Instructional Videos

The course teachers created a series of online IVs on a university-developed learning platform. These videos covered various topics related to business communication, such as interviewing, meetings, and product showcasing. The course comprised 10 units, with each

unit containing four 10-minute IVs.

3.2.2 Intelligent Tutoring System (iTutor)

iTutor, a university-developed AI-driven agent tool that is integrated into the course learning platform, is designed to enhance students' online learning experiences through AI-guided knowledge interaction(Xiao & Bai, 2022). The key interactive functionalities of iTutor that can aid in language learning were employed by the researchers to support the IVs. These functionalities include initiation, knowledge expansion, confirmation, evaluation and assessment, emotional support, and summarizing (Figure 1). Specifically, iTutor can initiate the learning process by posing questions or topics that guide students into conversation or activity. It can help students expand their language skills by introducing new vocabulary, grammatical structures, or cultural background information. Moreover, iTutor can confirm the correctness of learners' responses which boosts their confidence in their language abilities and evaluates their pronunciation by highlighting wrongly pronounced words. It can also provide motivational messages and supportive feedback to maintain student engagement and foster a positive learning experience. Further, it can summarize the key points discussed to reinforce the learning outcomes.

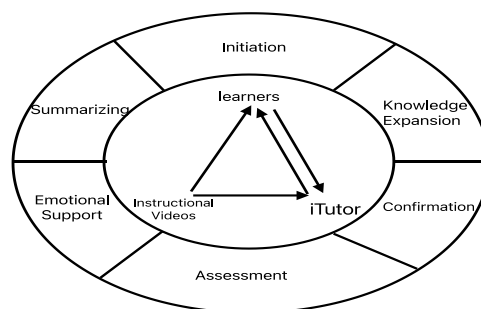


Figure 1. Design Framework of Instructional Videos with iTutor

The course teachers have tailored each lesson with iTutor, integrating questions and timed prompts to align with learning objectives and foster interactive video sessions. Figure 2 shows examples of the interface to realize iTutor's functions.

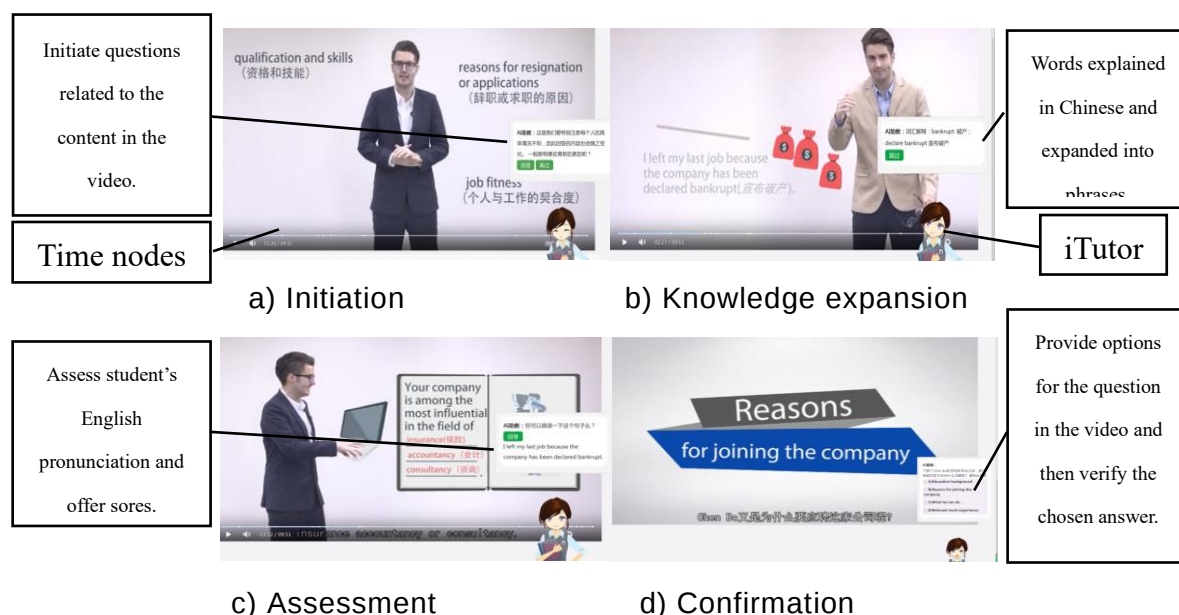


Figure 2. The interface of Instructional Videos with AI-driven iTutor

3.2.3 Language Assessment BEC Preliminary

In order to evaluate the listening and speaking skills of the learners, a pre- and post-test design was implemented. The test questions were taken from two mock tests of BEC Preliminary Test. The listening test consisted of four parts in the form of multiple-choice and blank-filling questions. The speaking test consisted of three parts in the form of short questions, mini-presentation, and discussion. The maximum scores for both the listening and speaking sections were converted into 100 points. The speaking test was rated by two experienced EFL teachers, using the scoring guides in BEC Preliminary Speaking. Inter-rater agreement for the two speaking tests was acceptable (pretest: $\alpha=.845$; posttest: $\alpha=.876$), and thus their average ratings were used to represent the learners' speaking scores.

3.2.4 Questionnaire

A questionnaire was used to gather data on the perceptions of learners towards the use of IV and IVIT. The questionnaire was adapted from Yang et al. (2022), who investigated perceptions towards AI chatbot language learning for English listening and speaking. The current study's questionnaire included 12 items to gather learners' perceptions from six aspects, including learning motivation, opportunity, anxiety, gain, overall learning experience, and willingness to use. Additionally, two open-ended questions were included in IVIT's questionnaire to investigate learners' perceived strengths and weaknesses of iTutor.

3.3 Procedure

Before the treatment, both groups took the mock BEC Preliminary listening and speaking pretest. After completing the pretest, the two groups were then given introductions to the IV

and IVIT. Then the learners of IV group were required to watch the videos while those of IVIT were required to watch the videos and interact with iTutor for ten weeks. Following the 10-week treatment, both groups completed the posttest and the questionnaire in the final week. Figure 3 illustrates the flow of the experiment.

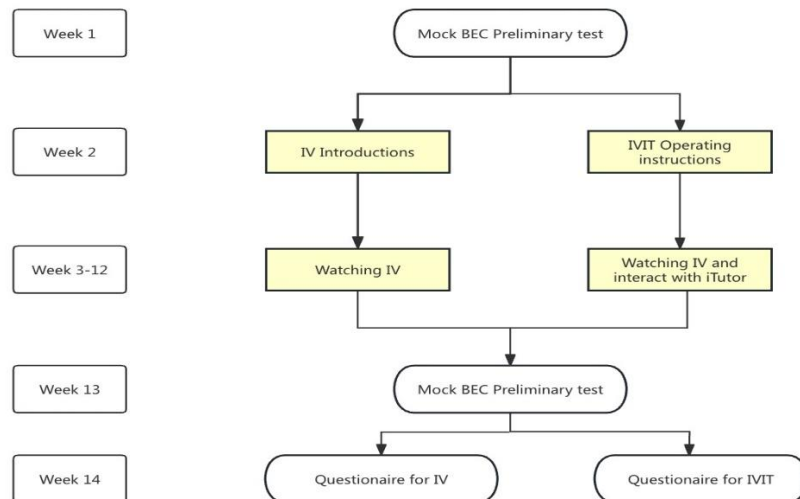


Figure 3. The Procedure of the Quasi-experiment

3.4 Data Analysis

The data analysis involved comparing the scores of two groups in pre-tests and post-tests, as well as their responses to a questionnaire. Paired sample t-tests were used to identify any differences within each group's pretest and posttest scores. One-way analysis of covariance (ANCOVA) was conducted to determine any differences in posttest scores between the two groups, while controlling for pretest scores. The Mann-Whitney U test was used to analyze the responses to the Likert-scaled questionnaire items to identify any differences in the learners' perceptions towards IV and IVIT. Additionally, the responses of group IVIT to the two open-ended questions were analyzed using a three-phase thematic analysis to identify the strengths, weaknesses or difficulties of IVIT.

4. Results

4.1 Learners' Listening and Speaking Development

The quantitative analysis involved the comparison of the pre-and post-test scores of the participants to measure improvements in listening and speaking skills. Table 1 indicates that learners from Group IV showed some improvement in listening and speaking skills, but it was not significant. In contrast, learners from Group IVIT showed significant advancements in both listening comprehension and speaking proficiency.

Table 1. Results of paired sample t-test on the pretests and the posttests

Group	Test	N	Mean (SD)		df	t-value	p	d
			Pre-test	Post-test				
IV	Listening	21	69.57(6.30)	71.77(7.50)	20	-1.852	0.079	0.404
	Speaking	21	72.69(5.42)	73.16(5.95)	20	-1.226	0.235	0.267
IVIT	Listening	22	74.60(8.60)	76.70(7.68)	21	-3.444	0.002**	0.734
	Speaking	22	71.03(6.84)	75.50(8.30)	21	-3.547	0.002**	0.756

* $p < .05$, ** $p < .01$, *** $p < .001$.

As shown in Table 2, one way Covariance (ANCOVA) was conducted to control for pre-test scores and revealed that IVIT group significantly outperformed the IV group in speaking proficiency ($p = .007$, $\eta^2 = .002$) in the post-test. However, there was no significant difference between the two groups in listening comprehension ($p = .598$, $\eta^2 = .0069$), suggesting that the use of iTutor had a more pronounced effect on the development of learners' speaking skills than their listening skills.

Table 2. Results of ANCOVA on post-test

Test	Group	N	Mean	SD	F	P	η^2
Listening	IV	21	71.77	7.50	0.283	0.598	0.002
	IVIT	22	76.70	7.68			
Speaking	IV	21	73.16	5.95	8.114	0.007**	0.069
	IVIT	22	75.50	8.30			

* $p < .05$, ** $p < .01$, *** $p < .001$.

4.2 Learners' Perceptions on IV and IVIT

4.2.1 Quantitative results

Table 3 presents a comprehensive analysis of learners' responses using IV/IVIT for their English language learning, measured through various constructs, i.e., Learning Motivation, Learning Opportunity, Learning Anxiety, Learning Gains, Overall Experiences, and Willingness to Use. As displayed in Table 3, learners from both groups were generally positive about IV and IVIT in facilitating language learning, showing no significant differences in their attitudes toward learning through IV and IVIT in terms of learning motivation, learning gains, overall experience, and willingness to use.

However, compared with Group IV, learners from Group IVIT viewed that iTutor could provide significantly more learning opportunities to improve speaking ability (Mean=4.18 vs. Mean=3.71, $p=0.03$). In addition, it is worth noting that learners from Group IVIT reported a significant more decrease in anxiety when speaking than those from Group IV (Mean = 4.18 vs. Mean = 3.52, $P=0.01$).

Table 3. Results of Mann-Whitney U test on questionnaire responses

Constructs	Item	IV		IVIT		U	p	r
		Mean	SD	Mean	SD			
Learning	1. Boost motivation for better English listening	4	0.55	4.09	0.29	249	0.53	0.54
Motivation	2. Boost motivation for better English listening	3.95	0.5	4.18	0.59	277	0.17	0.6
Learning	3. More learning opportunities to improve	3.86	0.85	3.91	0.75	232	0.99	0.5
Opportunity	English listening							
	4. More learning opportunities to improve speaking	3.71	0.72	4.18	0.59	310	0.03*	0.67
Learning	5. Nervous or anxious when listening	4.05	0.59	4	0.53	221.5	0.79	0.48
Anxiety	6. Nervous or anxious when speaking	3.52	0.98	4.18	0.5	321.5	0.01*	0.7
Learning	7. Improved listening after using it for self-	3.86	0.65	4	0.53	259.5	0.42	0.56
Gains	studying							
	8. Improved speaking after using it for self-	4.1	0.54	4.05	0.65	222.5	0.82	0.48
	studying							
Overall	9. Interesting and fun to use it for English self-	3.95	0.67	4.14	0.47	258.5	0.39	0.56
Experiences	studying							
	10. Interesting and fun to use it for English	3.86	0.57	4.05	0.58	267.5	0.29	0.58
	self-studying							
Willingness	11. Willing to continue using it for self-studying	3.76	0.62	3.95	0.58	269.5	0.21	0.58
to Use	English listening							
	12. Willing to continue using it for self-studying	3.81	0.68	3.82	0.66	232.5	0.98	0.5
	English speaking							

* $p < .05$, ** $p < .01$, *** $p < .001$. Scores of Item 5 and 6 were reversed to ensure consistency across all variables.

4.2.2 Qualitative Insights toward IVIT

Four themes concerning IVIT strengths emerged through qualitative analysis, including the interactive learning environment, real-time feedback, anxiety reduction, and accessibility. Learners mentioned that IVIT provided a dynamic platform where they could actively engage in speaking and listening simultaneously, which was more interactive compared to traditional video watching. They stated that active engagement was supported by features like bilingual support, allowing them to switch between languages to suit their comfort level. Moreover, learners reported that iTutor offered personalized diagnostic tools that provided immediate feedback on pronunciation and comprehension and helped them identify and focus on their weaknesses. In addition, IVIT was reported to reduce the stress typically associated with language learning in more formal settings by creating a more relaxing environment where learners did not fear judgment. Further, the learners pointed out that the system not only facilitated access to relevant videos and documents but also enhanced the overall learning process by recommending additional resources.

However, weaknesses of iTutor were also mentioned, including limited interaction and feedback, and technical constraints. The learners stated that sometimes iTutor struggled with complex queries and could not provide the nuanced feedback that a live instructor would,

limiting its effectiveness in deeper educational engagement. In using IVIT, the learners also faced technical issues such as audio re-recording difficulties, and challenges in accessing previous interactions, which could disrupt the learning flow and reduce usability.

5. Discussion

The first research question aimed to investigate the impact of two different language teaching methods, IV and IVIT, on the development of listening and speaking skills in EFL learners. Within-group statistical analysis indicated that Group IV did not exhibit significant improvements in either listening or speaking skills. In contrast, Group IVIT demonstrated substantial enhancements in both areas, with a notably higher progression in speaking skills compared to Group IV. This aligns with findings by Han (2020) and Shishido (2021), indicating that the use of AI-driven tools in the IVIT setup not only facilitated repeated practice but also significantly boosted learners' communicative confidence. However, Group IVIT did not outperform the Group IV in listening skills. A possible explanation for this finding is the short duration of exposure. The interaction with iTutor might be insufficient for listening substantial skill development, as long-term engagement is often necessary for listening proficiency (Yang et al., 2022).

The second research question explored learners' perceptions of the effectiveness of IVIT compared to traditional IV. Statistical analysis showed no significant differences in perceptions between the groups across ten evaluated items, with both groups generally responding positively. Notably, however, the IVIT group reported more frequent opportunities for speaking practice and experienced reduced anxiety during speaking tasks, which corroborates the quantitative findings of superior speaking performance in this group. These outcomes suggest that increased speaking opportunities and reduced anxiety are pivotal for EFL speaking skill development (Shishido, 2021).

Participants acknowledged the strengths of IVIT, echoing sentiments from Zou et al. (2023) and Han (2020) that highlighted the role of AI as a motivational tool that enhances learner engagement and mitigates anxiety, a significant obstacle in language acquisition. Despite these advantages, participants also reported challenges such as limited interaction with iTutor and recurring technical issues, which could hinder learning efficacy. These findings underscore the necessity for integrating AI systems into standard curricula carefully and evaluating their long-term impact on learning outcomes, as noted by VanLehn (2011) and Yüce et al. (2019).

In conclusion, while IVIT was found to enhance interactive learning paths and provide nuanced feedback effectively, the integration of AI-driven iTutor requires careful consideration to ensure it facilitates deeper interactions and addresses technical constraints for truly personalized learning experiences.

6. Conclusion

This study confirms that Interactive Video-based Instructional Technology (IVIT) has the potential to significantly improve the listening and speaking skills of adult English as a Foreign Language (EFL) learners. By comparing the outcomes of IVIT with traditional Instructional Video (IV), the study demonstrates the additional benefit that interactive elements bring to the educational process. Additionally, the overall positive learner perception highlights the willingness of adult language learners to embrace advanced AI-driven educational technologies. It would be useful to conduct further research to determine the varying effects of different types of interactive content and feedback mechanisms in IVIT on listening, speaking, and other EFL skills, which can provide valuable insights into the feasibility and long-term effectiveness of integrating IVIT into regular language curricula.

Acknowledgments

This study was supported by School of Humanities, Shanghai Open University (*Research on Online EFL Learners' Satisfaction*) and Shanghai Engineering Research Center of Open Distance Education (Grant NO. KEKT2008).

References

- Cinganotto, L., & Cuccurullo, D. (2015). The role of videos in the teaching and learning of content in a foreign language. *Journal of E-learning and Knowledge Society*, 11. <https://doi.org/10.20368/1971-8829/1024>
- Han, S. (2020). The effectiveness of voice-based AI chatbots in enhancing adult L2 learners' speaking skills. *Educational Technology Research and Development*, 68(4), 2035-2053. <https://doi.org/10.1007/s11423-020-09792-5>
- Kam, E. F., Liu, Y.-T., & Tseng, W.-T. (2020). Effects of modality preference and working memory capacity on captioned videos in enhancing L2 listening outcomes. *ReCALL*, 32(1), 213-230. <https://doi.org/10.1017/S0958344020000014>
- Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I., & Pineau, J. (2021). Automated data-driven generation of personalized pedagogical interventions in Intelligent Tutoring Systems. *International Journal of Artificial Intelligence in Education*, 32(2), 323-349. <https://doi.org/10.1007/s40593-021-00267-x>
- Kusumarasdyati. (2015). Incorporating culture into listening comprehension through presentation of movies. *TEFLIN Journal*, 16(1), 54-70. <https://doi.org/10.15639/teflinjournal.v16i1/54-70>
- Leiker, D., Ricker Gyllen, A., Eldesouky, I., & Cukurova, M. (2023). Generative AI for learning: Investigating the potential of synthetic learning videos. *ArXiv*. <https://doi.org/10.48550/arXiv.2304.03784>

- Ma, W., Adesope, O. O., Nesbit, J., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901-918. <https://doi.org/10.1037/A0037123>
- Mitrovic, A., Dimitrova, V., Lau, L., Weerasinghe, A., & Mathews, M. (2017). Supporting constructive video-based learning: Requirements elicitation from exploratory studies. *Educational Technology Research and Development*, 65, 991-1011. https://doi.org/10.1007/978-3-319-61425-0_1
- Saed, H., Haider, A. S., Al-Salman, S., & Hussein, R. (2021). The use of YouTube in developing the speaking skills of Jordanian EFL university students. *Heliyon*, 7. <https://doi.org/10.1016/j.heliyon.2021.e06317>
- Shishido, K. (2021). The role of technology in EFL education: Enhancing learning motivation and reducing anxiety. *TESOL Quarterly*, 55(3), 832-858. <https://doi.org/10.1002/tesq.548>
- Sun, Y.-C., & Yang, F.-Y. (2015). I help, therefore, I learn: service learning on Web 2.0 in an EFL speaking class. *Computer Assisted Language Learning*, 28(3), 202-219. <https://doi.org/10.1080/09588221.2013.818560>
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221. <https://doi.org/10.1080/00461520.2011.611369>
- Vungthong, S., Djonov, E., & Torr, J. (2017). Images as a resource for supporting vocabulary learning: a multimodal analysis of Thai EFL tablet apps for primary school children. *TESOL Quarterly*, 51, 32-58. <https://doi.org/10.1002/TESQ.274>
- Xiao, J., & Bai, Q. (2022). iTutor: Promoting AI guided knowledge interaction in online learning. 2022 *International Symposium on Educational Technology (ISET)*. <https://doi.org/10.1109/ISET55194.2022.00061>
- Yang, C. T.-Y., Lai, S.-L., & Chen, H. H.-J. (2022). The impact of intelligent personal assistants on learners' autonomous learning of second language listening and speaking. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2022.2141266>
- Yüce, A., Abubakar, A., & Ilkan, M. (2019). Intelligent tutoring systems and learning performance. *Online Information Review*, 43(4), 600-616. <https://doi.org/10.1108/OIR-11-2017-0340>
- Yükselir, C., & Kömür, Ş. (2017). Using online videos to improve speaking abilities of EFL learners. *European Journal of Education Studies*, 3, 255-266. <https://doi.org/10.6084/EJES.V0I0.654>
- Zou, H., Tan, S. P., & Lee, J. H. (2023). Adult EFL learners' perception of AI-driven speech evaluation programs as tools for English conversation and pronunciation improvement. *Journal of Language Learning Technologies*, 34(1), 45-62. <https://doi.org/10.1002/llt.3032>