AI in Language Education: Enhancing Learners' Speaking Awareness through AI-Supported Training

Fan Yang[1,*], Kangxi Li[1], and Ruilin Li[2]

[1]School of Foreign Studies, Hefei University of Technology, Hefei, China
[2]Institute of Higher Education, Dalian University of Technology, Dalian, China
Email: F.YangEdu@foxmail.com (F.Y.); richardlkx@126.com (K.L.); lirl21@dlut.edu.cn (R.L.)
[*]Corresponding author
Manuscript received February 5, 2024; revised March 12, 2024; accepted March 25, 2024; published June 17, 2024

***Abstract*—This study reported teaching practices using AI-supported training to enhance English language learners' speaking awareness in instructional contexts. The study involved 36 intermediate-level English language learners in engineering majors from a research university in China. Participants engaged in a language training project using a spoken dialog system called TalkAI as technical support. The training process and its impact were recorded and analyzed with quantitative data derived from AI scoring feedback on pronunciation, grammar, and usage performance, as well as qualitative data obtained from student self-reports. Results indicate that students made significant progress in overall speaking performance through AI-powered training, as evidenced by improvements in scoring reports and the length of student responses. In addition, students managed to articulate specific statements regarding their speaking skills, develop systematic judgments concerning personal performances, and deploy targeted strategies in formulating follow-up action plans. These findings underscore the potential of AI-supported training in fostering positive shifts in students' language awareness within instructional contexts.**

***Keywords*—language awareness, generative AI, scoring rubrics, rubric training, speech skills, educational technology**

I. Introduction

Although the importance of educational technology in facilitating language education has long been recognized, the advent of ChatGPT has spurred a wide range of discussions on how the evolution of Artificial Intelligence (AI), especially Generative AI (GAI), might influence language learning and education. For researchers, the interactive flexibility, immediacy, and accuracy of AI have opened new opportunities not only for learners but also for teachers to build a personalized language learning community where diverse learning needs could be satisfied and teaching efficiency could be improved. This community can satisfy diverse learning needs and improve teaching efficiency, potentially sparking a wave of educational reform. However, existing studies in language education have predominantly focused on the broader impact of AI on the language learning ecosystem or individual perceptions of AI education, overlooking the considerable potential of generative AI as an effective teaching tool within language classrooms [1, 2]. To address this gap, our study investigated student responses from an awareness-raising perspective within an AI-supported language training project.

This study investigates and analyzes the potential of AI in language education by reporting innovative teaching practices that incorporate an AI-powered Spoken Dialog System (SDS) named TalkAI into the speaking assessment process. TalkAI is an online spoken dialog system that could interact with users through conversation, offering additional assessment and feedback services on users' English language performance. Drawing on the training approaches of Glover [3] and Su [4] for enhancing language awareness, data of student responses to the SDS system, especially their changes in speaking performance and awareness, were collected.

In this study, we were particularly interested in exploring student responses to AI-supported training from a language awareness perspective, as well as the pedagogical implications of these factors for both language educators and learners. By reporting innovative teaching practices that incorporate an AI-powered educational platform into the speaking assessment process within instructional contexts, this study investigates and analyzes the potential of AI in language education, viewing it not only as a supplementary teaching tool but also as an assessment coordinator and research assistant, aiming to relieve teachers from heavy marking loads and to provide more opportunities for research and reflection in language classrooms.

II. Literature Review

A. Language Awareness

Language Awareness (LA), defined as the "explicit knowledge about language and conscious perception and sensitivity in language learning, language teaching and language use" [5], has long been acknowledged as a crucial factor in facilitating learning progress through "means of cognitive, meta-cognitive, and metalinguistic awareness and strategies" [3, 6, 7]. The integration of the LA approach within the domain of language learning, however, was not prominent until the 1980s. Prior to that, language awareness primarily existed as a broad concept and belief among language teachers, largely based on observational insights within classroom settings [8]. Later, the importance and validity of LA were further substantiated through empirical studies focusing on the relationship between bilingual children's phonological awareness and their language proficiency [8, 9]. This led to the recognition of language awareness as a pivotal pedagogical movement, significantly influencing the research field of language learning and teaching.

Recent decades have seen extensive discussions about the term Language Awareness (LA) among practitioners, theorists, and researchers in various educational settings. Theorists primarily focus on identifying effective teaching approaches to LA and its relationship to self-assessment. Meanwhile, empirical studies have delved into LA's possible

impact on language skills, notably in areas of speaking and reading.

In the realm of teaching and training approaches, scholars adhering to the traditional behaviorist perspective contend that LA should be viewed as the behavioral outcomes of a gradual awareness-raising process. They advocate for a structured, repetitive approach in LA training, focusing on incremental knowledge acquisition in areas such as pronunciation and vocabulary [10]. In contrast, cognitive linguists perceive LA as a critical meta-cognitive construct with an "affective" element, as it "engages and helps to evolve attitudes and values" [11], and therefore LA training should prioritize fostering learners' self-assessment capabilities and encouraging critical reflection on their learning experiences. This concept of reflective learning is further endorsed by communicative linguists, who propose that LA can assist learners in understanding and evaluating their linguistic and cultural identities in communication [12].

The emphasis on self-assessment within LA was then extended to the broader realms of assessment and evaluation. Educators and policymakers began to weave LA principles into language teaching curricula, including reference materials and proficiency descriptors. A notable instance of of this integration is observed in the Common European Framework of Reference (CEFR), renowned for its can-do-statement style level-specific descriptor [13]. These can-do statements, as highlighted in relevant studies, not only serves as acknowledged indicators of language proficiency, but are also designed as useful tools for self-assessment, enabling students to evaluate their own language abilities [3].

In terms of empirical research, previous LA research primarily concentrated on its impact on relatively easy-to-test sub-skills such as listening and reading, productive skills including speaking and writing, especially speaking, have received less attention. Among the studies that do address these areas, one notable investigation on speaking skills demonstrated that LA integration enables students to provide "longer, more relevant, and more detailed and critical descriptions of their speaking skills" [3]. In the context of writing, several studies have focused more on the pedagogical side, highlighting how LA can aid teachers and examiners in standardizing the scoring scales and more effectively assessing students writing tasks [14]. Overall, the incorporation of LA in teaching and training is increasingly recognized as an essential component in language education.

B. LA and Speaking Assessment

As previously mentioned, much of the past research on Language Awareness (LA) has incorporated LA training within the assessment process. This incorporation typically involves integrating [14], analyzing [4], or even co-constructing assessment rubrics [3], serving as the main tools and benchmarks for both teaching and evaluating LA. Specifically, in the development of rubrics for assessing speaking skills, common measurement tasks including group discussion [15], structured interview [16], and individual response to short prompts [17] often employ the four-dimensional Complexity, Accuracy, Lexis, and Fluency (CALF) approach [18] in test design, guiding both assessors and test-takers to focus on key areas of speaking performance, namely the grammatical structure, lexical richness, content relevance, and pronunciation clarity.

In addition to focusing on various assessment dimensions, LA training might also equip students with the skills to accurately recognize and articulate different performance levels [3, 4]. To facilitate this, researchers often utilize official language proficiency frameworks, including the Common European Framework of Reference (CEFR), the Canadian Language Benchmarks (CLB), and China's Standards of English Language Ability (CSE) in the development of assessment rubrics [19]. These frameworks provide level-specific descriptors, exemplified by can-do-statements that describe specific language proficiency features at different levels. For instance, a CSE speaking descriptor at level 5 states, "Can, after preparation, briefly comment on topics in his/her field". With clear can-do-statements as reference, researchers can devise rubric training that helps students set explicit learning targets, cultivate systematic language awareness, and consequently enhance their speaking abilities.

C. AI and Speaking Assessment

In contrast to the multi-dimensional and level-specific rubrics used by human assessors, AI-based automated speaking assessment systems predominantly focus on identifying specific language features for scoring purposes. These systems meticulously extract various features such as pronunciation, fluency, vocabulary, grammar, and semantics from the language recordings of test takers. Then these extracted elements are assessed using a deep neural network model [20] or automated speech scoring methods such as multiple regression and classification trees [21, 22]. This methodology of feature extraction and measurement is also adopted by generative AI-powered online speaking programs, such as TalkAI. These programs often include sophisticated scoring modules as part of their services offerings to attract potential users.

Recent research has also delved into spoken interactions elicited from AI-mediated interaction and the application of automatic speech recognition technology for teaching and assessment purposes [22–24]. The use of Spoken Dialog Systems (SDS) and Intelligent Personal Assistants (IPAs), such as Alexa have been noted for their potential to enhance interactive opportunities and engagement in educational settings. Despite these advancements, the validity and reliability of these technological tools as a replacement for human assessors were still not testified. Earlier studies have identified limitations in AI-supported SDS systems, particularly their inability to fully comprehend users input in from prior dialogue turns, resulting in unnatural conversational flow and therefore challenges in accurate scoring [25].

D. LA Training

To enhance language awareness and performance, researchers have implemented various assessment practices rooted in different rubric training approaches, namely lecture-based [26] and practice-oriented [3] training. The first approach places emphasis on teacher-centered lecturing and sampling, where the teacher thoroughly explains each dimension of the assessment rubrics, using samples as demonstration [4]. The second practice-oriented approach

shifts the focus towards student-centered activities, requiring students to actively participate in task design, rubric development, peer-assessment, standardization, and self-assessment procedures [3]. Proponents of the lectured-based approach argue that it establishes firm assessment criteria for students who act as assessors, while advocates of the practice-oriented approach reported increased student engagement in their hands-on practices [3, 26]. More recently, the SIB (Studying rubrics, Identifying performance features, Bridging the gap) rubric training model had been introduced and tested. This model has been recognized as particularly effective in the process of raising awareness in language education, suggesting a new direction in rubric-focused teaching methodologies [4].

However, all previously reported practices were conducted in human-assessor contexts; none investigated the potential benefits of combining speaking rubric training with AI support to enhance learners' speaking awareness and skills. To address these gaps, this study focuses on exploring student responses to AI-supported training in authentic English speaking classrooms, examining the training from the perspective of language awareness. The study seeks to answer the following research questions:

RQ1. Were there any changes in student's AI scoring performance before and after the training? If so, in what aspects?

RQ2. Were there any changes in student's speaking awareness before and after the training? If so, in what aspects?

III. Methodology

A. Participants

Participants in this study consisted of 36 intermediate-level English language learners from a research university in China. Previous empirical studies usually included sample sizes ranging from 30 to 65 participants (with 32 in [3] and 62 in [4], typically involving one or two teaching classes receiving similar training input), guiding the decision for this study to invite one class of students to participate in the training. In the sample class, 36 out of 39 participants managed to submit all required tasks throughout the training. All participants were year 1 engineering students learning English as a Foreign Language (EFL), aged between 17 and 19 years. The participants was predominantly male, with 27 male students, and 9 female students. Entry exam data indicated that most learners started at level 5–6 of China's Standards of English Language Ability in terms of general English proficiency.

B. TalkAI

The AI speaking program used in this study was TalkAI (https://ttalkai.com/), an AI-powered Spoken Dialogue System (SDS) capable of engaging users in conversations on selected or spontaneous topics. Additionally, it provides scoring and feedback services (refer to Fig. 1) on users' English language performance across three dimensions: pronunciation, grammar, and content. Employing advanced AI-powered voice analytics, TalkAI provides also offers detailed pronunciation score breakdowns (see Fig. 1) on fluency, accuracy and prosody. Its scoring engine is based on the extraction and calculation of targeted spoken language features pertinent to the dimensions mentioned above.

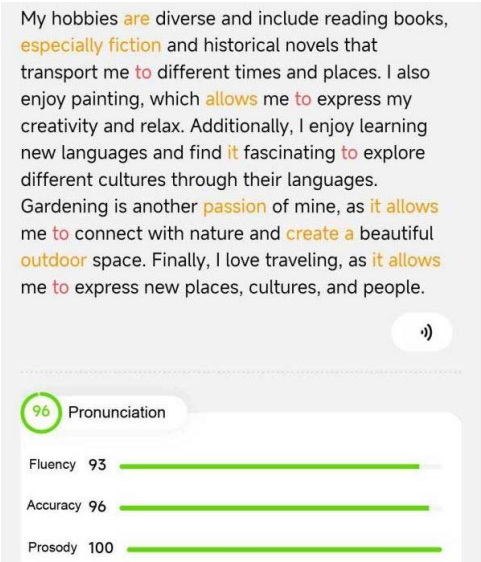

Fig. 1. TalkAI feedback sample.

C. Procedure

Rubric training was conducted over one teaching semester at the university from November to December 2023. It comprised eight weekly teaching sessions, with the training of pronunciation, grammar and content each assigned two sessions, and pre- and first and final sessions dedicated to pre- and post training assessments. The training followed a modified version of the SIB (Studying rubrics, Identifying performance features, Bridging the gap) training cycle as proposed by [2], with an additional reflection (R) section added for collecting student reports. Details of the training scheme are illustrated in Fig. 2:

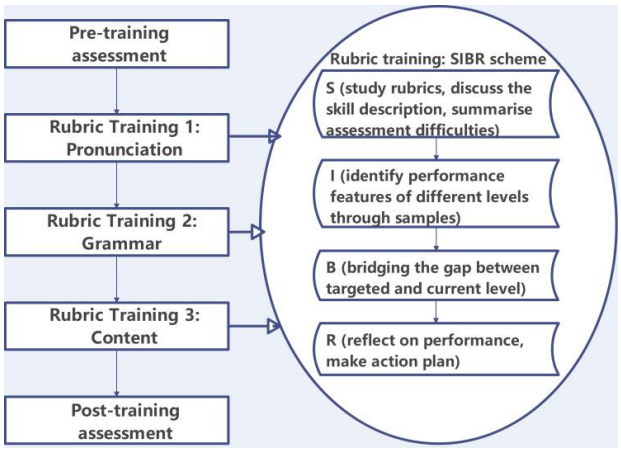

Fig. 2. Rubric training scheme.

D. Data Collection and Analysis

Data was collected during teaching sessions. In both pre- and post-training assessment sessions, students were engaged in interactions with TalkAI on a consistent conversational topic, responding to at least two follow-up questions. Recordings from these AI interactions, along with scoring reports and feedback, were collected for further analysis. Additionally, participants compiled self-reports in which they were encouraged to freely comment on their personal speaking performance and impression of the AI tool.

The quantitative data derived from AI scoring reports on students' speaking performance were descriptively analyzed and compared from three key dimensions: pronunciation,

grammar, and content (usage). For the qualitative data from self-reports, a three-phase coding process was employed, as per [27]: open coding, axial coding, and selective coding. Reports were first coded and then classified into broader categories. For example, a comment like "I had a few pauses" was initially coded as "fluency" in the first phase, and then as "pronunciation" in the second phase, and then as a specific awareness-raising sample in the final selective coding phase.

IV. Results

A. Changes in Speaking Proficiency

Results from the pre- and post-study recordings, corroborated by AI scoring reports, indicated potential improvements in students' speaking proficiency. In both assessments phases, TalkAI generated scoring reports for participants' spoken dialogue performances across three dimensions: pronunciation, grammar, and usage and content, with integer scores ranging from 0 to 100 in each dimension. A significant increase in scores across all three dimensions was observed after the training, as evidenced by paired-sample t-tests (see Table 1). Descriptive data revealed that the most notable improvement was in pronunciation scores ($t = 3.051$, $p = 0.004$), followed by grammar ($t = 2.528$, $p = 0.016$) and content ($t = 2.305$, $p = 0.027$). The varying effect sizes across these dimensions could stem from students' different speaking practice preferences or, more likely, from inherent scoring biases in the GAI system. Previous studies have suggested that automated language evaluation systems often provide more accurate scoring and feedback in areas like grammar and speech recognition, in comparison to occasional failure in content evaluation [28]. This assumption also found supportive evidence in self-reports, with several students mentioning easier improvements in pronunciation and grammar, in contrast to content scores, which "didn't grow as expected, even after several attempts".

Table 1. Descriptive statistics of scoring report

Dimensions	Pre-study		Post-study		Paired-sample t-test	
	M	SD	M	SD	t	p
Pronunciation	87.75	17.13	95.44	4.74	3.051	0.004*
Grammar	90.50	6.93	93.67	2.65	2.528	0.016*
Content	86.92	8.21	90.25	2.80	2.305	0.027*

Note. M = mean; SD = standard deviation; * $p < 0.05$.

The analysis of recording data also revealed an increase in the duration of students' spoken dialogues with TalkAI. The average length of these interactions grew from 65 seconds to 88 seconds, suggesting that students were more inclined to articulate extensive content following their exposure to rubric training. Such an enhancement in speaking duration could be indicative of improved confidence and a deeper engagement in speaking practices, potentially facilitated by the structured approach of the rubric training.

B. Changes in Speaking Awareness

Analysis of pre- and post-study self-reports disclosed that students acknowledged the role of AI-supported training in enhancing their language learning awareness from diverse perspectives. Several notable changes in their awareness have been identified:

Contrasting the self-reports collected pre-study and post-study revealed a marked increase in detail and specificity. Initially, the average word count of reports was 63 words; however, in final reports, this average had risen to 167 words. The expanded length of the reports allowed students to describe their speaking skills with more specific and relevant details. Whereas pre-study comments tended to be general and brief, post-study reflections demonstrated a shift towards specificity and depth. For example, one student initially noted, "I can use some advanced language" in his first report. In final reports, he included a more nuanced self-assessment: "I think my answer is native-like and natural with some complex word usage like collocations and linking words".

The AI-supported training also enabled students to develop more systematic judgments of their speaking skills. In pre-study reports, students tended to rely on surface-level reasons to explain and assess their speaking performance. Common explanations included statements such as "I had been learning mute English in high school", "I haven't received systematic training on oral English", and "My spoken language foundation is weak". After training, they began making critical evaluations from various dimensions in self-reports. They used more accurate descriptors to describe their spoken language, such as "my language is very fluent, without any pause or self-repetition", "I have some occasional pauses and wrong pronunciation in my recording", and "I also used many clauses in the article, such as ellipsis, attributive clauses and so on". These examples showed how students learned to evaluate specific sub-skills such as pronunciation and grammar, and that they could employ level-specific indicators including "occasional pause" or "without any pause" to systematically judge their proficiency levels.

An additional noteworthy change was in students' approach to learning strategies. The first obvious change was the shift from a reason-oriented approach to a solution-oriented approach in their assessment practices. Initially, many students in their pre-study reports focused on explaining their shortcoming in achieving satisfactory scores from TalkAI, often attributing these to factors such as their language background or learning habits. In contrast, the final reports reflected a more proactive attitude. Students began proposing detailed action plans to enhance their speaking proficiency. These included specific strategies like "writing down topic words" and "working on my chunking". In addition, certain reports highlighted effective ways of using TalkAI as GAI speaking assistant, including "paying attention to the words marked in red and yellow on your transcript", and "making full use of AI's suggestions on your content". These responses indicate a transition towards the internalization of learning strategies and a more constructive use of AI tools in language learning.

V. Discussion

This study explored the effectiveness of AI-supported training in enhancing the speaking awareness of English language learners. A comparative analysis of data gathered before and after the training sessions indicated a marked improvement in students' scoring performance. In addition, students showed positive changes in the awareness-raising training process, which was evident in the increased accuracy

of their self-assessments and overall judgments, and the adoption of more effective and suitable learning strategies consistently reflected in self-reports. The results from this study thus provide robust empirical evidence, contributing significantly to the ongoing scholarly discourse in the field of language education technology.

A. Rubric Training: Supported by AI and Supporting AI

This study demonstrated that the rubric training process in teaching contexts can be supported by AI. With teacher guidance, students were able to engage comprehensively with the training through studying the AI marking dimensions and grading reports, identifying different performance features from instant feedback, bridging performance gaps, and conducting in-depth self-reflection. The flexibility and immediacy of AI scoring and feedback service not only relieves teacher-assessors from marking load in rubric training, but also empower students to identify personal mistakes and set targeted goals through meaningful interaction with the system.

The findings of this study also highlighted the necessity of training support for effective integration of AI in educational contexts. Although the release of ChatGPT has injected a wide range of discussions in the educational field, it was observed that many students used GAI tools merely as search engines for homework assistance. Rubric training, however, facilitates a more comprehensive engagement with AI tools. It encourages student-users to consistently interact with, and critically reflect on their use of these tools. Such regular practice and reflection not only enhance their learning experience but also potentially contribute to a feedback repository. This repository can be invaluable for AI developers, offering insights to refine and advance their products.

B. Raising Awareness: Controllable and Beneficial for Engineering Students

The awareness-raising results in this study proved that language awareness can indeed be augmented through targeted classroom interventions. Furthermore, it was observed that increased awareness might lead to positive changes in scoring performance, a significant indicator of enhanced language proficiency. This adds advantageous evidence to previous research results in language sub-skills of reading and interpretation, testifying that the sub-skill of speaking can also significantly benefit from a structured awareness-raising process.

In addition, the inclusion of engineering students as participants in the study proved that language awareness interventions might benefit not only English majors or language-focused disciplines, but also engineering majors. The empirical evidence gathered from engineering students provides proof that language awareness is not only applicable but also advantageous for individuals pursuing different academic trajectories. The results underscore the universality of language awareness as a foundational skill with far-reaching implications beyond traditional linguistic domains, implying that language proficiency should be considered as as a crucial asset in interdisciplinary academic pursuits and professional endeavors.

C. AI Education in the New Era: Changing Roles of Various Classroom Participants

This study also urged educators and teachers to reconsider the roles of different classroom participants in the era of generative AI. The modern classrooms have witnessed the changes in roles of students becoming centers and teachers becoming information providers, facilitators, and assessors in the teaching process. In this context, the study introduced AI as a primary source of training and assessment, and it effectively assumed these roles. In AI-supported classrooms, the dynamic interplay among teachers, students, AI, and educational institutions necessitates careful examination. It's crucial to analyze both the evolving and newly emerged educational roles of these stakeholders and to explore their implications thoroughly. Such understanding is essential to optimize the use of AI in educational settings and to navigate the shifting landscape of teaching and learning.

D. Limitation

This study also has its limitations. A primary constraint was the reliance on AI-generated scoring reports for analyzing changes in students' speaking performance, without incorporating external evaluations from human-assessors. In addition, the methodology involved collecting self-reports and scores only at two points (before and after training), neglecting the fact that awareness-raising was a continual process. To address these limitations, future research could benefit from a more longitudinal approach. This could include multiple assessments throughout the training period, and incorporating evaluations from both GAI and human-assessors. Such a comprehensive approach would provide a more nuanced understanding of the learning trajectory and offer richer, more longitudinal evidence of language development.

VI. Conclusion

This study reported teaching practices using AI-supported training to enhance the speaking awareness of English language learners within instructional contexts. A comparative analysis of data collected before and after the training revealed significant improvements in AI scoring reports, suggesting enhanced speaking performance among learners. After training, participants managed to use more relevant and specific statements to describe their speaking skills. They also developed systematic approaches to evaluating their performances and adopted more targeted strategies for language learning, indicating a positive increase in language awareness. The findings of this study underscore the potential of AI in language education, as well as the role of modern classrooms in facilitating its widespread adoption. It is therefore hoped that this study will inspire various stakeholders including teachers, learners, and educational institutions to actively engage in and support the ongoing "AI plus" educational reform tide.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The authors collectively contributed to all aspects of this

research endeavor. F.Y. and K.L. designed and conducted the research over one teaching semester, while F.Y. and R.L. were responsible for the data analysis. They all contributed to the composition of the paper and they were actively involved in revising the manuscript. All authors reviewed and approved the final version of the manuscript, confirming their agreement with the findings and conclusions presented.

FUNDING

This work was supported by China Foreign Language Education Fund (Grant no. ZGWYJYJJ11A038).

REFERENCES

[1] N. Hockly, "Artificial intelligence in English language teaching: The good, the bad and the ugly," *Relc Journal*, vol. 54, no. 2, pp. 445–451, 2023.

[2] M. H. Al-khresheh, "Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching," *Computers and Education: Artificial Intelligence*, 2024.

[3] P. Glover, "Using CEFR level descriptors to raise university students' awareness of their speaking skills," *Language Awareness*, vol. 20, no. 2, pp. 121–133, 2011.

[4] W. Su, "Exploring how rubric training influences students' assessment and awareness of interpreting," *Language Awareness*, vol. 29, no. 2, pp. 178–196, 2020.

[5] Association for Language Awareness (ALA), "Language awareness defined," 2009.

[6] J. Cummins, "Critical multilingual language awareness: The role of teachers as language activists and knowledge generators," *Language Awareness*, vol. 32, no. 4, pp. 560–573, 2023.

[7] R. Ellis, "Principles of instructed language learning," *System*, vol. 33, pp. 209–224, 2005.

[8] R. Carter, "Language awareness," *ELT Journal*, vol. 57, no. 1, pp. 64–65, 2003.

[9] E. Bialystok and J. Codd, "Cardinal limits: Evidence from language awareness and bilingualism for developing concepts of number," *Cognitive Development*, vol. 12, no. 1, pp. 85–106, 1997.

[10] R. P. Leow, "Attention, awareness, and foreign language behavior," *Language Learning*, vol. 47, no. 3, pp. 467–505, 1997.

[11] T. Wright and R. Bolitho, "Language awareness: a missing link in language teacher education?" *ELT Journal*, vol. 47, no. 4, pp. 292–304, 1993.

[12] L. K. Barany, "Language awareness, intercultural awareness and communicative language teaching: Towards language education," *International Journal of Humanities and Cultural Studies (IJHCS)*, vol. 2, no. 4, pp. 257–282, 2016.

[13] C. Jones and R. Carter, "Literature and language awareness: Using literature to achieve CEFR outcomes," *Journal of Second Language Teaching & Research*, vol. 1, no. 1, pp. 69–82, 2012.

[14] J. Pun and K. K. C. Cheung, "Developing and validating a genre awareness questionnaire for writing scientific reports," *Frontiers in Psychology*, vol. 14, 1210240, 2023.

[15] S. Nematizadeh and Y. Cao, "Investigating willingness to communicate in synchronous group discussion tasks: one step closer towards authentic communication," *International Review of Applied Linguistics in Language Teaching*, 2023.

[16] S. J. Ross, *Interviewing for language proficiency: Interaction and interpretation*, Palgrave Macmillan, 2017.

[17] X. Qiu, "Picture or non-picture? The influence of narrative task types on lower-and higher-proficiency EFL learners' oral production," *International Review of Applied Linguistics in Language Teaching*, vol. 60, no. 2, pp. 383–409, 2022.

[18] F. Pang and P. Skehan, "Performance profiles on second language speaking tasks," *The Modern Language Journal*, vol. 105, no. 1, pp. 371–390, 2021.

[19] J. Liu, "China's standards of English language ability and English teaching," *Foreign Language World*, no. 3, pp. 7–14, 2019.

[20] Y. Wang, "Detecting pronunciation errors in spoken English tests based on multifeature fusion algorithm," *Complexity*, 2021, pp. 1–11.

[21] X. Xi, D. Higgins, K. Zechner, and D. Williamson, "A comparison of two scoring methods for an automated speech scoring system," *Language Testing*, vol. 29, no. 3, pp. 371–394, 2012.

[22] T. A. Sharadgah and R. A. Sa'di, "A systematic review of research on the use of artificial intelligence in English language teaching and learning (2015-2021): What are the current effects?" *Journal of Information Technology Education: Research*, vol. 21, 2022.

[23] T. M. Phillips, A. Saleh, and G. Ozogul, "An AI toolkit to support teacher reflection," *International Journal of Artificial Intelligence in Education*, vol. 33, no. 3, pp. 635–658, 2023.

[24] S. Jung Youn, "Test design and validity evidence of interactive speaking assessment in the era of emerging technologies," *Language Testing*, vol. 40, no. 1, pp. 54–60, 2023.

[25] G. J. Ockey and E. Chukharev-Hudilainen, "Human versus computer partner in the paired oral discussion test," *Applied Linguistics*, vol. 42, no. 5, pp. 924–944, 2021.

[26] M. Patri, "The influence of peer feedback on self-and peer-assessment of oral skills," *Language Testing*, vol. 19, no. 2, pp. 109–131, 2002.

[27] A. L. Strauss and J. M. Corbin, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 2nd ed., Thousand Oaks, CA, USA: Sage Publications, 1998.

[28] L. Qian, Y. Yang, and Y. Zhao, "Syntactic complexity revisited: sensitivity of China's AES-generated scores to syntactic measures, effects of discourse-mode and topic," *Reading and Writing*, vol. 34, no. 3, pp. 681–704, 2021.

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).