Digital Inclusivity In EFL Oracy Development: Implementing The Multimodal Cognitive Oracy Model For Arabic Learners In Online Higher Education

Received 2025-06-16
Accepted 2026-09-19
Published 2026-03-19

V S Sreelakshmi, S Vijayakumar, M Ahamedullah, A Sathikulameen
[1,2,3]B.S. Abdur Rahman Crescent Institute of Science and Technology, India, [4]The New College, Chennai, India
sreelakshmi_english_jan2024@crescent.education[1], vijayakumar@crescent.education[2*], ahamedullah@crescent.education[3], drsathikulameen@gmail.com[4]

To cite this article: Sreelakshmi, V S., Vijayakumar, S., Ahamedullah, M., Sathikulameen, A. (2026). Digital Inclusivity In EFL Oracy Development: Implementing The Multimodal Cognitive Oracy Model For Arabic Learners In Online Higher Education. Ijaz Arabi: Journal of Arabic Learning, 9 (2), 648-666, DOI: https://doi.org/10.18860 /ijazarabi.V9i2.39744

Abstract

Inclusive education in English as a Foreign Language (EFL) has become increasingly critical in online higher education, particularly for Arabic-speaking learners who face unique linguistic challenges in developing oracy skills. This study evaluates the effectiveness of the Multimodal Cognitive Oracy Model (MCOM) in enhancing listening and speaking competencies among Arabic learners at B.S. Abdur Rahman Crescent Institute of Science and Technology, School of Arabic and Islamic Studies. A mixed-methods quasi-experimental design was employed with Arabic-speaking EFL learners enrolled in online degree programs. Participants were divided into experimental and control groups, with the experimental group receiving instruction through the MCOM framework incorporating computer-assisted language learning (CALL) tools, multimodal inputs (audio, visual, textual, kinesthetic), and cognitive scaffolding strategies. The eight-week intervention integrated authentic English materials, synchronous and asynchronous interactive activities, and formative feedback mechanisms. Data collection included pre- and post-intervention assessments of listening comprehension and speaking proficiency, supplemented by qualitative feedback through questionnaires and reflective journals. Results demonstrated statistically significant improvements in both listening comprehension and speaking fluency for the experimental group. Participants exhibited enhanced phonological awareness, improved pronunciation accuracy, increased oral communication confidence, and better comprehension of varied English accents. Qualitative analysis revealed high learner engagement and satisfaction with multimodal activities. Visual and auditory scaffolding effectively addressed L1 interference and facilitated cognitive processing of English phonemes. The study confirms that the Multimodal Cognitive Oracy Model provides an effective, inclusive pedagogical framework for Arabic-speaking EFL learners in online contexts. Practical implications emphasise curriculum integration of technology-enhanced oracy activities, educator training in CALL methodologies, and institutional digital infrastructure investment. Future research should explore longitudinal impacts and cross-cultural applications of the MCOM framework.

Keywords: Digital; EFL; Oracy Development; Multimodal Cognitive Oracy Model; Arabic-Speaking Learners; Online; Higher Education

INTRODUCTION

Higher education has experienced a radical change in inclusion into a pedagogical model, particularly in teaching English as a Foreign Language (EFL). Inclusive education confirms that all learners, irrespective of their lingual backgrounds, cultural orientations, and modes of learning, have the right to have equal access to quality language learning opportunities. This principle is especially important in the case of Arabic-speaking learners in online higher education because of profound phonological, morphological, and syntactic differences between Arabic and English. The process of building strong oracy capabilities, the skill to talk clearly and grammatically and knowing how to comprehend various spoken messages, is much more difficult to develop in an entirely online setting without direct feedback, where social interactions and multimodal feedback are missing. Arabic speakers encounter unique language challenges in acquiring the English language. English has phonemes that are not found in Arabic (e.g. /p t k/ and some vowel differences) and has phonemic rhythm (of stress), uncharacteristic of the Arabic syllable-based one. Such differences are the causes of continued pronunciation mistakes, problems in understanding speech, and decreased confidence in speaking. Furthermore, the consonantal root system of Arabic and derivational morphology contribute to developing cognitive processing patterns, which are not easily transferred to English, which requires new linguistic strategies (Bashori et al., 2024).

Historically, the focus of traditional EFL instruction has tended to focus on literacy, reading, and writing, with little or no attention paid to systematic oracy development. Such an imbalance leaves most learners of the Arabic language unprepared to communicate in oral skills in real life, although they have studied it formally for years. This gap has become more urgent and potential with the rapid growth of online learning, which is to be fueled by the COVID-19 pandemic (Al Fraidan and Alaliwi, 2024). Computer-Assisted Language Learning (CALL) promises good solutions. Interactive, adaptive and multimodal language practice is supported with the help of CALL technologies (automatic speech recognition system (ASR), chatbots with AI, real-world multimedia materials, and collaborative platforms). These instruments will offer the Arabic learners multiple repetitive exposures to target sounds, the ability to develop the skills individually, alleviate anxiety, due to the opportunities to practice privately, and engage with various types of English and natural speech (Sun, 2023; Chiu et al., 2024).

Yet significant gaps remain. A lot of CALL literature is based on generic EFL populations. It does not take into account the needs of Arabic speakers nor explores the tools in a fully integrated and theory-oriented framework. In higher education via the online realm, where students need to go through sophisticated online worlds to develop highly developed language skills, an inclusive model of coherence is indispensable. The study is reactive to this by proposing and analysing the Multimodal Cognitive Oracy Model (MCOM) as a detailed model that follows an online situation, with Arabic-speaking EFL learners. MCOM combines the theory of Cognitive Load Theory, Multimodal Learning Theory and evidence-based CALL practices to progressively formulate the skills of listening and speaking. This study, as conducted in B.S. Abdur Rahman Crescent Institute of Science and Technology, provides a significant contribution to the research area, i.e., theoretical understanding of inclusive and technology-enhanced EFL theoretical pedagogy as well as practical advice in online learning settings.

DOI: 10.18860 /ijazarabi.V9i2.39744
ISSN(print): 2620-5912 | ISSN(online):2620-5947
ejournal.uin-malang.ac.id/index.php/ijazarabi/index

Computer-Assisted Language Learning (CALL) has also seen the development of drill-based activities to communicative and interactive environments that approximate the use of the language in natural settings, and the positive yet inconsistent outcomes of the depth of the effect on EFL oracy development are continuously reported (Yang and Chen, 2007; Ghanizadeh et al., 2015). One of the major discoveries is Automatic Speech Recognition (ASR) technology that provides non-judgmental and instant feedback on pronunciation and fluency to acquire learner autonomy and decrease the fear of speaking (McCrocklin, 2016). Meta-analysis results validate medium to large pronunciation intelligibility effects of ASR that rely on the levels of proficiency, duration of intervention, and targets (Ngo et al., 2024). Sun (2023) also demonstrated that ASR not only increase fluency and accuracy, but also metacognitive awareness about oral production. In the case of Arabic-speaking learners, ASR is particularly useful in helping learners overcome the L1-L2 phonological gaps that include /p/-/b/ opposition or untypical vowel characteristics with the help of the phoneme-level feedback. Jiang et al. (2023) and Bashori et al. (2024) determined through their studies that ASR-integrated tasks were particularly helpful in increasing the accuracy and fluency of speaking, along with the learner's confidence in speaking.

In addition to pronunciation, CALL helps to facilitate oracy based on meaning in the form of a meaningful interaction that is built on the concepts of SLA, such as negotiation of meaning, comprehensible input, and pushed output (Bailey et al., 2021). Digital platforms promote interactions between learners and computers (e.g., chatbots), learners and learners (e.g., video conferencing), and between learners and content (e.g., multimedia). Synchronous ones enhance real-time fluency and pragmatics, whereas the asynchronous format permits reflexive, revising speech (Takase, 2024). The emergence of AI-based chatbots is becoming a key factor, and it provides learners who are online and speak Arabic with a low anxiety level an opportunity to train flexibility with adaptive dialogues. AbuSahyon et al. (2023) emphasise that they promote motivation and practice in speaking, and Chiu et al. (2024) demonstrate that the desired perceived usefulness and autonomy support predict the continuity of participation. This is essential in environments with limited exposure to native speakers. Lastly, technology-based language learning, eventually, Mulyadi et al. (2021) found that ESP students improved their listening and speaking; yet, they appreciated the nature and the intent of online activities. Zhou and Wei (2018) go further to note that effective students integrate metacognitive, cognitive, and socioaffective approaches when applying digital technologies, highlighting the importance of explicit teaching of strategies alongside technology adoption.

The foundation of the Multimodal Cognitive Oracy Model (MCOM) is the multimodal learning theory, whereby learning is enhanced when presented by multiple senses as opposed to one sense. Gilakjani et al. (2011) demonstrated that this simultaneous input contributes to better memory of praise and learning among different personality styles, which is congruent with the assertion of Dual Coding Theory that the combination of both verbal and visual processing results in enhanced cognitive representations. This method is especially familiar to digital environments that incorporate the natural integration of text, audio, video, animation and interactivity. Satar (2015) proposed a set of principles that should be used to design multimodal affordances to facilitate learner interaction during online language interactions, and Pellicer-Sanchez (2022) revealed that the inclusion of text, images, and audio in learning language reading materials results in higher comprehension and vocabulary acquisition rates as compared

DOI: 10.18860 /ijazarabi.V9i2.39744
ISSN(print): 2620-5912 | ISSN(online):2620-5947

to text-only designs. These also wear over to oracy: multimodal materials, e.g. audio with transcripts and visual aids, are useful to help students comprehend speech and develop an insight into English sound-spelling connections and discourse patterns.

In the case of Arabic-speaking EFL learners, multimodal scaffolding is important since there are large orthographic and phonological differences between Arabic and English. Such visual aids as phonetic transcription, articulatory diagrams and animated intonation models play a role in filling the (relatively) low L1-to-L2 transfer in pronunciation. According to Giannakos and Cukurova (2023), successful multimodal teaching needs a close correspondence of theory, design and assessment to formulate modalities as being usable and complementary to each other, removing cognitive overload and optimising the learning process. Cognitive Load Theory (CLT) is an alternative theory of optimisation of online oracy teaching. CLT distinguishes between intrinsic load (difficulty of tasks inherent in them, and aggravated by L1-L2 distance), extraneous load (caused by ill design) and germane load (effort used in schema building). Typical differences in the mastery of Arabic by the learners lead to high intrinsic load, but, when assisted by strategically designed schedules to help the students focus, would help achieve a low extrinsic load (Ozer and Kilic, 2018). The recent research confirms it in digital models: Tatli et al. (2022) have established that well-coordinated multimedia in digital storytelling enhanced speaking and vocabulary and reduced perceived cognitive load; Yu et al. (2019) have established that coherent navigation and learner control of mobile systems decreased extraneous load without interfering with the engagement; and Ma (2025) has established expectation confirmation and perceived usefulness as the mediating variables of how EFL students experience cognitive demands in MOOCs. Given the simultaneous demands of decoding unfamiliar phonology, accessing incomplete lexical representations, constructing non-native syntax, and managing digital interfaces, Arabic-speaking learners in online settings especially benefit from CLT-informed design. By minimising extraneous load, instruction can free cognitive resources for the germane processing essential to oracy development.

Digital inclusivity goes beyond technological access to open and equitable spaces of meaningful participation, authentic engagement, and successful learning outcomes for all students, regardless of background, ability, or condition. For the Arabic-speaking EFL learners attending online higher education, these include linguistic barriers, cultural norms of content and interaction, technology gaps, and educational strategies that are responsive to a wide range of learning contexts (Hopkyns, 2023). While digital spaces are promising tools for democratising education, recent research has warned that they may reproduce or worsen inequalities without inclusive considerations. In a case study conducted at the Arab Open University, Beloufa (2023) found that virtual spaces often favour Western communication practices and native English skills, while disadvantaging Arabic-speaking learners. This research study emphasises culturally responsive education to validate multiple epistemologies and communicative styles and to accelerate development toward goal language proficiency.

Language itself is a critical axis of digital inclusivity. Malkawi et al. (2024) demonstrated that Arabic-speaking users expend significantly more time and cognitive effort than native English speakers to achieve comparable outcomes on English-dominant digital platforms, underscoring the need for inclusive interface design, intuitive navigation, and linguistically accessible instructional language. Nguyen et al. (2022) similarly support a move away from native speakers of English in digital higher education

by creating an international language that can accommodate multiple speakers and speech-making uses.

The intersection of digital inclusivity and oracy creates both challenges and opportunities. Au et al. (2024) found that Arabic-speaking older Canadians could only progress digitally through language learning when platforms offered multilingual interfaces, culturally-appropriate content, and activity structures consistent with the understandings from earlier learning that apply equally as well to current students in classes. Thus, effective online oracy instruction must address challenging language as well as culturally aware scaffolding for full participation. Although Afril et al. (2024) focus on the study of Arabic, they propose an understanding of what successful digital language instruction must be, which is that it links L1 aspirations to L2 goal attainment, provides multiple developmental pathways and supports meaningful community practice. These findings indicate that inclusive EFL oracy instruction is dependent upon more than technological knowledge, but on intentional, learner-centred instruction.

The current findings on CALL, multimodal learning, and digital inclusivity have remained fundamental in supporting oracy practice among Arabic-speaking EFL learners in online higher education. Most of these studies will use individual tools or specific instructional practices rather than developing comprehensive, theoretically integrated systems. On top of that, Arabic-speaking learners who struggle with a specific linguistic distance from English are less prevalent than the general EFL population or speakers of typologically closer languages. This rapid growth in online instruction has further narrowed the gap between practice and evidence-based oracy instruction in fully digital settings. MCOM addresses these gaps by combining Cognitive Load Theory, Multimodal Learning Theory and evidence-based CALL practices into a single model for Arabic learners. This study, conducted and evaluated at B.S. Abdur Rahman Crescent Institute of Science and Technology, provides theoretical foundations for effective technology-based oracy design and examples for engaging with digital inclusion in EFL education.

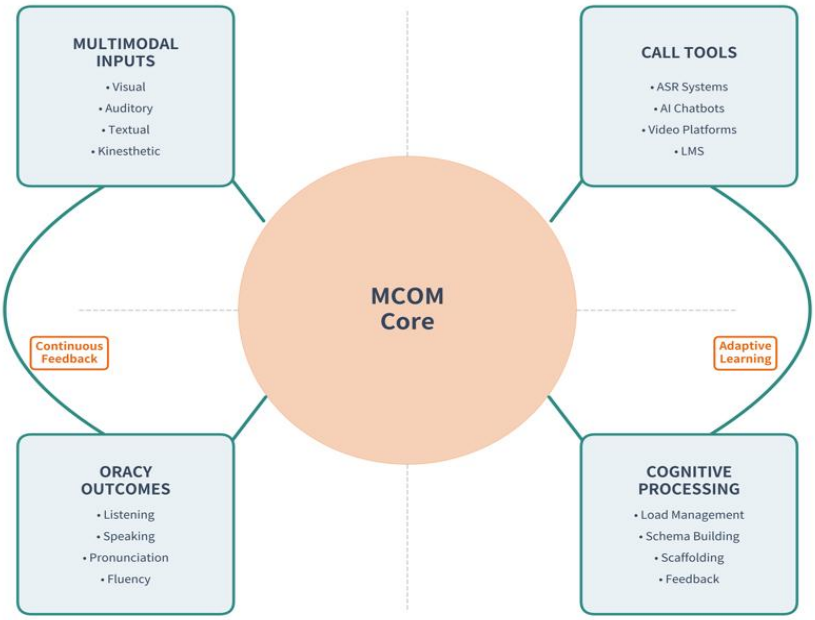
The Multimodal Cognitive Oracy Model (MCOM) is an all-encompassing pedagogic model that addresses the oracy developmental needs of the EFL Arabic-speaking learners in online higher education. It combines Cognitive Load Theory (CLT), Multimodal Learning Theory, and evidence-based CALL practices into a single approach with the capacity to inform the process of instructional design, technology choice, sequence of activities and assessment. Fundamentally, MCOM uses CLT to maximise the limited cognitive resources of the learners. There is a high level of intrinsic cognitive load on Arabic-speaking learners processing English because there are great differences in phonological, morphological and syntactic differences between L1 and L2. In response to this, MCOM uses design techniques that reduce extraneous load and encourage germane processing, such as sequencing tasks progressively, attention to important linguistic features, breaking down complex tasks, and scaffolded practice, which is gradually faded away as proficiency improves.

The Multimodal Learning Theory is also integrated in MCOM based on the idea of using several different sensory modalities to support oracy development. In listening, the learners work on audio and visual aids, including transcripts, concept images, and discourse structure graphic organisers. For speaking, they rely on written cues, visual cues and audio samples and are able to write, reread and edit their output. Such multimodal scaffolds provide a remedy to the constraints of L1 transfer, especially in English phonology, by rendering abstract features more tangible using visual and kinesthetic

DOI: 10.18860 /ijazarabi.V9i2.39744
ISSN(print): 2620-5912 | ISSN(online):2620-5947
ejournal.uin-malang.ac.id/index.php/ijazarabi/index

means. As Figure 1 illustrates, MCOM divides these principles into a four-component system (Auditory Input Enhancement, Phonological Development, Interactive Oral Practice, and Metacognitive Development), which is based on CLT and Multimodal Learning Theory, mediated by CALL technologies, and oriented to digital inclusion. The development of oracy is therefore constructed as a cyclic process, which interrelates with each other, and with the theory, pedagogy and technology mutually reinforcing each other.

Figure 1. The Multimodal Cognitive Oracy Model (MCOM), CALL Technology, Digital Inclusivity Orientation



The MCOM operationalises its theoretical foundations through four interrelated instructional components, each supported by targeted CALL technologies.

1. Auditory Input Enhancement helps learners develop listening understanding with incremental exposure to authentic English speech in all accents, speaking rates and discourse types. This component contains curated multimedia libraries, podcast services, and interactive listening apps like EdPuzzle that allow students to control playback speed, access transcripts and be provided with embedded comprehension aids. These activities progress from highly scaffolded, such as advance organisers, segmented input, to more naturalistic listening experiences that respond to real-world needs.
2. Phonological Development addresses pronunciation issues specific to Arabic speakers by delineating segments (i.e., /p/vs. /b/) and suprasegmental features (stress, rhythm, intonation). CALL tools include ASR systems such as ELSA Speak for phoneme-level feedback, visual displays of pitch/intonation, comparisons with native models on both sides, and focused drills. This multimodal approach draws on auditory representations, visual representations of sound, kinesthetic awareness of the articulatory sense, and corrective response in the moment to make sound-like abstract features apparent.
3. Interactive Oral Practice provides a variety of opportunities for meaningful spoken production. AI chatbots for low anxiety conversations like Replika, synchronous

video conferencing for peer interaction like Zoom breakout rooms, and synchronous platforms such as Flipgrid for reflective speaking with revision, and simulation communication. Activity sequence moves from form-based, structured tasks towards open-ended, fluent tasks that prioritise pragmatic relevance over accuracy.

4. Metacognitive Development fosters learning in awareness and self-regulation through digital portfolios that contain the recordings of samples and reflection, LMS-based analysis of participation and progress, and guided reflection prompts. It is a component that prepares the student for lifelong oracy development outside of class by providing strategic, independent learning habits.

These components operate cyclically, reinforcing one another within a unified framework that integrates cognitive, linguistic, and metacognitive dimensions of oracy. Implementation follows a structured sequence: needs analysis → scaffolded skill-building with embedded formative assessment → culminating performance tasks in authentic communicative contexts. Throughout, the model prioritises cognitive load management, multimodal scaffolding, and inclusive design responsive to diverse learner profiles. Based on the review and findings, the following research questions were formulated.

This study describes what is the effect of MCOM-based instruction on the listening comprehension and speaking proficiency of Arabic-speaking EFL learners in online higher education? To what extent do different components of the MCOM (auditory input enhancement, phonological development, interactive oral practice, and metacognitive development) contribute to observed improvements in oracy skills? How do Arabic-speaking learners perceive the effectiveness, usability, and inclusivity of MCOM-based instruction for their oracy development? What implementation challenges and facilitating factors emerge during the deployment of the MCOM in authentic online higher education contexts?

METHOD

This study employed a mixed-methods quasi-experimental design to investigate the effectiveness of the Multimodal Cognitive Oracy Model for Arabic-speaking EFL learners in online higher education. The quasi-experimental approach was selected because random assignment of students to experimental and control conditions was not feasible within the institutional context; instead, intact class sections were designated as experimental or control groups. The mixed-methods design of listening comprehension and speaking skills using quantitative data on the perceptions and experiences of students provides a holistic view of the learning outcomes and processes that produce those outcomes.

The independent variable was the instructional design, where the experimental group received instruction in MCOM with four components: auditory input enhancement, phonological development, interactive oral practice, and metacognitive development, all supported by one or more CALL technologies. The control group received traditional online EFL instruction as outlined in the institutional standard curriculum, which included listening and speaking but without any systematic use of CALL tools or the theoretical framework guiding the MCOM approach. Dependent variables included listening comprehension measured through standardised and curriculum-aligned assessment activities; speaking proficiency assessed through holistic and analytic rating scales for sampled voice; and perception of the learner measured through questionnaires, interviews

and reflective journals. Other variables of interest were technology acceptance, cognitive load and learner engagement, which were assessed to establish factors that mediate the connection between instruction and learning outcomes.

The sample consisted of 68 EFL Arabic-speaking learners who were undergraduate degree students at the B.S. Abdur Rahman Crescent Institute of Science and Technology, School of Arabic and Islamic Studies. All participants were native Arabic speakers who had attended a minimum of 6 years of primary and secondary education, studying English as a foreign language. The sample was between 18 and 22 years old. They were the representatives of the different regional backgrounds of the Arabic-speaking world (students of India (Tamil Nadu), Saudi Arabia, the United Arab Emirates) and the other Gulf states. The experimental group had 34 participants who were enrolled in two units of an online course on English communications. In contrast, the control group had 34 participants enrolled in two comparison units of an online course on the same topic. Baseline testing ensured that there were no significant group differences on such pre-intervention measures as English proficiency, prior English learning experience, technological competence, or demographic variables.

All instruction occurred entirely online due to institutional delivery formats for this program. The online learning platform employed the use of a learning management system (Moodle) as the main content delivery, management of activities and communication medium. Synchronous sessions were done through Zoom video conferencing, and the sessions were to be held twice a week, with a duration of 90 minutes. Learners did asynchronous activities, such as self-paced listening practice, pronunciation exercises and reflective tasks, between synchronous sessions, independently. The experimental group intervention based on MCOM was carried out during a period of eight weeks, and activities were sequenced well so as to develop oracy skills over time. There were two 90-minute synchronous meetings and a total of about three hours of asynchronous learning activities every week. The four integrated elements of the MCOM model were included in the intervention, and each element was supported with particular CALL technologies. The Auditory Input Enhancement activities were done with a collection of diverse high-quality audio and video material in English representing a variety of topics, accents, and discourse types. Resources were TED talks, BBC learning English podcasts, audio excerpts of academic lectures, and dialogue discussions. The learners were exposed to such materials using guided listening tasks, which supported pre-listening vocabulary, in-listening comprehension checks with access to transcripts and post-listening reflection activities. The EdPuzzle platform allowed embedding interactive questions into video materials, and learners were allowed to speed up playback and transcripts to facilitate learning.

Phonological Development tasks were aimed at addressing particular pronunciation issues of Arabic speakers, such as problematic consonant contrasts, vowel quality differences, word stress patterns, and sentence-level intonation. ELSA Speak mobile application offered ASR-based feedback about the accuracy of pronunciation, and students were required to go through 10-minute daily practice sessions involving algorithmically determined pronunciation priorities. The synchronous classroom sessions incorporated direct teaching of the phonology of the English language with the help of visual representation, e.g. vowel charts, articulatory charts and spectrogram displays of the variation between the learner and target forms. Praat software allowed learners to visualise and analyse patterns, pitch and stress in their recorded speech in comparison to

native speaker patterns. Interactive Oral Practice involved human-human interaction mode and human-computer interaction mode. The structured speaking activities involved in synchronous sessions, such as information-gap tasks, problem-solving discussions and role-plays, involved meaningful oral interaction of learners. Small groups of conversation were facilitated by breakout rooms, and the instructor went through the groups to give feedback and assist. The Replika AI chatbot was used to complement human interaction, where the learners got a chance to engage in long-term dialogue based on self-selected topics, and the chatbot responded to their conversation, giving them a chance to keep the conversation alive with the chatbot occasionally modelling the target structures. Flipgrid was an asynchronous speaking platform where students would make video-based responses to prompts and give peer feedback on the recordings of their classmates, forming a non-synchronous community of practice that would continue to exist even after the conclusion of synchronous classroom time.

Metacognitive Development activities helped the learners to be aware of the processes of learning and their strategic competency. Every week, the learners answered structured reflection prompts, answering questions like: What did you focus on regarding pronunciation this week? What strategies helped you understand challenging listening materials? How has your confidence in oral communication changed? What goals do you have for next week's practice? These reflections were documented in digital portfolios that also housed recorded speech samples collected at multiple time points throughout the intervention. Periodic review of portfolio contents enabled learners to identify patterns of growth and areas requiring continued attention. The control group experienced conventional online EFL instruction that included listening and speaking activities, but without the systematic CALL integration or theoretical framework characterising the MCOM approach. Control group instruction followed the institution's standard textbook-based curriculum, with listening activities based on textbook audio materials and speaking practice occurring primarily through teacher-led question-and-answer exchanges in synchronous sessions. While control group learners had access to the same learning management system and video conferencing tools as the experimental group, they did not engage with the specialised CALL applications (ELSA Speak, EdPuzzle, Replika, Flipgrid) or receive the systematic phonological instruction and multimodal scaffolding central to the MCOM intervention.

Data collection occurred at multiple time points using diverse instruments to address the study's research questions comprehensively. Quantitative data on listening comprehension and speaking proficiency were collected through pre-intervention, immediate post-intervention, and delayed post-intervention assessments (administered two weeks after the intervention concluded). Qualitative data capturing learner perceptions and experiences were collected continuously throughout the intervention period and through focused inquiry at its conclusion. Listening comprehension was assessed using a locally developed instrument aligned with the course learning outcomes and validated through expert review and pilot testing. The assessment included three sections: short conversation comprehension with multiple-choice questions testing literal and inferential comprehension, extended listening passages (academic lecture excerpts and narrative descriptions) with open-ended comprehension questions, and dictation tasks requiring accurate transcription of short sentences containing phonological features challenging for Arabic speakers. The scale was marked on a holistic basis with a possible total of 100, and inter-rater reliability was set at 0.89 by doubling the scoring of a subset

of assessments. Speaking proficiency was assessed using recorded speech samples and was provoked by four types of tasks: a read-aloud passage to check pronunciation accuracy and fluency under the condition of a controlled conditions, picture description requiring spontaneous speech in regard to visual stimuli, personal narrative prompt, learners were asked to describe a meaningful experience, and conversational role-play simulating a real communicational situation. Two trained assessors rated the recordings through an analytic rubric based on the IELTS speaking band descriptors, with different scores on pronunciation, fluency, coherence, lexical resource, and grammatical range and accuracy. The analytic scores were added together to form a total speaking proficiency score (maximum 100 points). Inter-rater reliability calculated through intraclass correlation coefficients was found to be above 0.85 in all sub-dimensions.

The data were collected using various methods in order to achieve triangulation and to fully understand the experiences of the learners. At the end of the intervention, all the members of the experimental group were given a Likert-scale questionnaire that dealt with perceptions of instructional effectiveness, technology usability, cognitive load, and learning outcomes (n=34). The questionnaire contained closed-ended questions, as well as open-ended questions that used elaborated answers. Semi-structured interviews were utilised with a purposive sample of 12 respondents in the experimental group who would have varied levels of proficiency and patterns of engagement. Interviews addressed the perception of the participants on the aspects of the intervention that benefited them the most, the challenges faced and the strategies used, as well as the areas where the intervention can be improved. The interviews were all audio taped, transcribed and thematically analysed. Last, the reflective journal entries made weekly by learners during the intervention offered longitudinal qualitative data on the changing perceptions and experiences.

RESULTS AND DISCUSSION

Quantitative data analysis answered research questions about how MCOM-based instruction has influenced listening comprehension and speaking proficiency and the relative influences of various components of interventions. Preliminary analyses examined data on assumptions of parametric statistical tests such as normality, equal variance and independent observations. Measurement of participants and outcome variables was done through descriptive statistics (means, standard deviations, frequency distributions). To establish the presence of significant differences in the results of listening comprehension and speaking proficiency between experimental and control groups, mixed-design ANOVAs were performed using group (experimental vs. control) as a between-subject factor, and time (pre-intervention, immediate post-intervention, delayed post-intervention) as a within-subject factor. Greater group-by-time effects would demonstrate that groups are growing differently and the MCOM intervention is effective. The size of the effect was estimated by partial eta squared (ηp^2) in order to estimate the magnitude of the observed effect. Post-hoc comparisons with Bonferroni corrections were used to determine time points at which the groups differ significantly.

In order to analyse the relative importance of various components of MCOM to generate oracy, multiple regression analyses were performed with data from only experimental groups. Predictor variables consisted of the measures of interaction in every component of an intervention: hours spent in auditory input activities, the number of pronunciation practice sessions done, the amount of interactive oral practice (chatbot

conversations and peer interactions), and the amount of reflective journal entries made. Post-intervention and pre-intervention changes in the outcome variables were to achieve scores in listening comprehension and speaking proficiency. Standardised regression coefficients (β) revealed the relative value of each component in the prediction of outcomes, taking into account the rest of the predictors. The qualitative data analysis followed the process of thematic coding, which was done repeatedly to determine the patterns of learner perceptions and experiences. NVivo qualitative data analysis software was used in analysing questionnaire open-ended responses, interview transcripts, and journal entries. Primary coding used descriptive codes of content on the surface. Analytical coding was later conducted, and related codes were packed into themes, which are like recurring motifs in the data sources. The ultimate thematic categories were developed after undergoing constant comparison processes, and member checked with participants of the interviews. Combination of quantitative and qualitative results was at the interpretative stage, where the qualitative themes were used as explanations of the quantitative trends.

Ethical Considerations

De-identification was done to keep participant data confidential. All data files used pseudonyms rather than real names of participants, with the master list of pseudonyms and identities kept separately in a secure spot accessible only to the research team. Reported findings present only aggregate data.

The first two research questions on the efficacy of MCOM-based instruction in listening comprehension and speaking proficiency were quantified. Descriptive statistics and inferential testing proved significant advantages for the experimental group in the two outcome domains.

Table 1. Descriptive Statistics for Listening Comprehension Scores

Group	Pre-test M (SD)	Post-test M (SD)	Delayed Post-test M (SD)	Gain M (SD)
Exp (n=34)	64.8 (8.2)	78.5 (7.1)	76.9 (7.4)	13.7 (6.8)
Cont (n=34)	65.2 (8.5)	69.8 (8.3)	69.3 (8.1)	4.6 (4.9)

Note. M = Mean; SD = Standard Deviation. Scores range from 0 to 100.

As in Table 1, baseline performance on pre-test measures was comparable for both groups, but no significant differences were observed between the experimental and control groups ($t(66) = 0.19$, $p = .851$). But there was a considerable discrepancy at post-test and delayed post-test assessment points. The mean increase in listening comprehension in the experimental sample was 13.7 points, an increase of about 21% from baseline. The difference was modest, with 4.6 points, or 7% improvement over the control group. Mixed-design ANOVA revealed a significant group-by-time interaction effect ($F(2, 132) = 18.43$, $p < .001$, $\eta p^2 = 0.22$), confirming differential growth patterns between groups.

Table 2. Descriptive Statistics for Speaking Proficiency Scores by Dimension

Group	Dimension	Pre-test M (SD)	Post-test M (SD)	Gain M (SD)
Experimental	Overall	62.3 (9.7)	74.8 (8.2)	12.5 (6.3)
	Pronunciation	14.2 (2.8)	18.3 (2.1)	4.1 (2.2)
	Fluency/Coherence	15.8 (2.9)	18.9 (2.4)	3.1 (2.1)
	Lexical Resource	16.1 (2.7)	18.7 (2.3)	2.6 (1.9)
	Grammatical Range	16.2 (2.8)	18.9 (2.5)	2.7 (2.0)
Control	Overall	63.1 (9.4)	66.8 (9.1)	3.7 (4.2)
	Pronunciation	14.5 (2.7)	15.4 (2.6)	0.9 (1.6)
	Fluency/Coherence	15.9 (2.8)	16.7 (2.7)	0.8 (1.5)

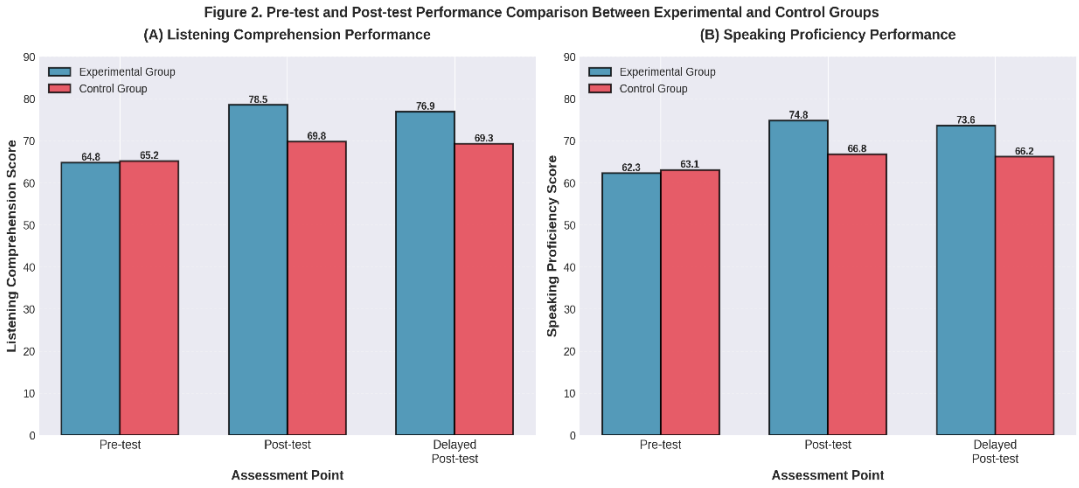
DOI: 10.18860 /ijazarabi.V9i2.39744
ISSN(print): 2620-5912 | ISSN(online):2620-5947
ejournal.uin-malang.ac.id/index.php/ijazarabi/index

	Lexical Resource	16.3 (2.6)	17.1 (2.5)	0.8 (1.4)
	Grammatical Range	16.4 (2.9)	17.6 (2.8)	1.2 (1.7)

Note. M = Mean; SD = Standard Deviation. Overall scores range from 0-100; sub-dimension scores range from 0-25 each.

Table 2 presents speaking proficiency outcomes disaggregated by assessment dimension. Analysis of overall speaking proficiency scores using mixed-design ANOVA revealed a significant group-by-time interaction ($F(1, 66) = 32.17$, $p < .001$, $\eta p^2 = 0.33$), indicating differential improvement patterns between groups. The experimental group achieved a mean gain of 12.5 points (20% improvement), compared to 3.7 points (6% improvement) for the control group. Examination of speaking sub-dimensions revealed that experimental group advantages were most pronounced for pronunciation (gain difference = 3.2 points, $t(66) = 6.89$, $p < .001$, $d = 1.68$) and fluency/coherence (gain difference = 2.3 points, $t(66) = 5.12$, $p < .001$, $d = 1.24$), with more modest advantages for lexical resource and grammatical range. These patterns coincide with the explicit emphasis on phonological development and oral practice during the MCOM intervention. A visual representation of trends in performance across assessment points for the experimental and control groups is shown in Figure 2, which provides substantial and sustained gains for the learners who received MCOM instruction.

Figure 2. Pre-Test, Post-Test, and Delayed Post-Test Performance



To address the second research question of how MCOM components interact in different ways, multiple regressions were performed on experimental group data. These predictor variables included time spent on each intervention component, number of hours spent on auditory input, number of pronunciation practice sessions completed, number of oral interactions (chatbot conversations, peer interactions), and number of reflective journal entries submitted. Table 3 reports the standardised regression coefficients for listening comprehension and speaking proficiency gain scores.

Table 3. Multiple Regression Results: MCOM Component Contributions to Oracy Gains

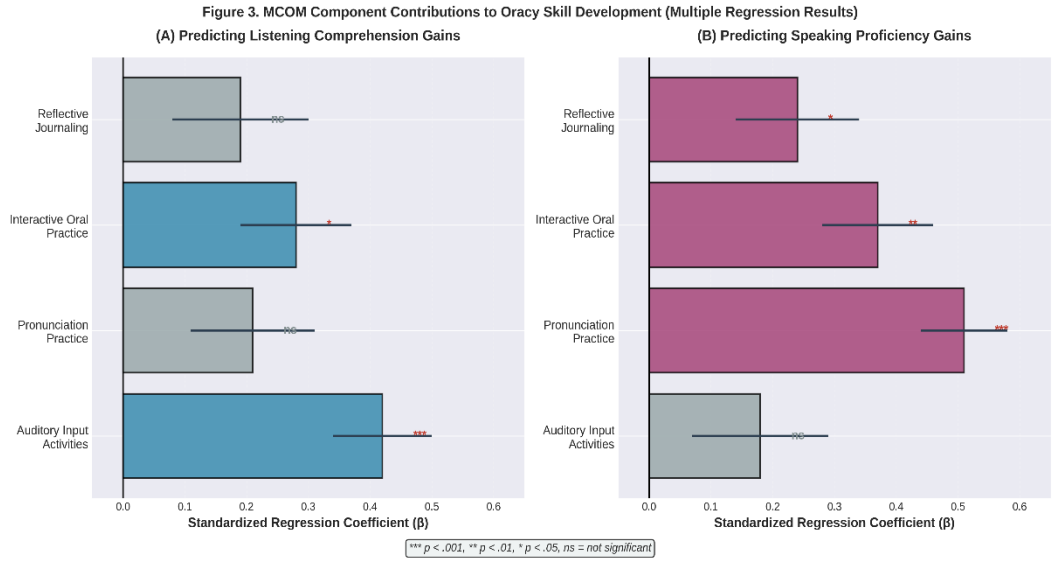
Predictor Variable	Listening Gain β	p-value	Speaking Gain β	p-value
Auditory Input Activities (hours)	0.42	.002**	0.18	.187
Pronunciation Practice (sessions)	0.21	.098	0.51	<.001***
Interactive Oral Practice (quantity)	0.28	.024*	0.37	.004**
Reflective Journaling (entries)	0.19	.121	0.24	.048*
Model R^2	0.58	,	0.67	,

Note. β = standardised regression coefficient. *** p < .001, ** p < .01, * p < .05.

DOI: 10.18860 /ijazarabi.V9i2.39744
ISSN(print): 2620-5912 | ISSN(online):2620-5947
ejournal.uin-malang.ac.id/index.php/ijazarabi/index

To develop listening comprehension improvement, activities involving auditory input were found to be the best predictor (β = 0.42, p =.002), which is theoretically valid since such activities were focused on the direct development of listening skills. Interactive oral practice was also an important predictor of the advancements in listening (β= 0.28, p =.024), indicating that the activities related to oral production were a contributor to better comprehension, possibly because of a better phonological representation and better overall confidence with oral language. Pronunciation practice sessions were most predictive (β = 0.51, p <.001), indicating the direct focus on speaking skills with the help of ASR-based feedback and explicit instruction in phonological skills. Interactive oral practice was also a major predictor of speaking gains (β = 0.37, p =.004), which validated the significance of genuine production chances. Interestingly, the relationship between reflective journaling and speaking gains (β = 0.24, p =.048) was also shown to be small but significant, indicating that metacognitional level aided the gain of skills. The visual display of these regression findings is given in Figure 3. It shows that the variable contributions of MCOM components in contributing to listening and speaking outcomes are different, and the error bars show how accurate estimates can be.

Figure 3. *Standardised regression coefficients (β) showing the relative contributions of MCOM components to (A) listening comprehension gains and (B) speaking proficiency gains. Error bars represent standard errors. Significance levels: *** p < .001, ** p < .01, * p < .05, ns = not significant. Colored bars indicate statistically significant predictors; grey bars indicate non-significant predictors.*



Qualitative Results: Learner Perceptions and Experiences

Research using questionnaire responses, interviews, and journal entries identified four themes indicating the students' perceptions of the MCOM intervention: (1) they had increased confidence and reduced anxiety; (2) they appreciated multimodal and technology supports; (3) they appreciated individualised pace and repeated practice; and (4) they appreciated an authentic communication focus. Second theme was Recognition of Multimodal and Technological Supports; learners had a particular interest in the use of visual, auditory, and interactive features. Participants frequently described the use of transcript access for listening, visual representations of pronunciation, and the ability to record, review, and revise their speech: "The combination of hearing the audio, seeing the transcript, and having questions embedded in the video helped me understand much

better than just listening alone" (Participant E23, questionnaire). For example, a participant described pronunciation practice as "Seeing the visual representation of my intonation on the model helped me understand what I needed to change." "Just hearing was not enough" (Participant E14, interview). These arguments support the multimodal design concept that has been central to the MCOM framework. Third was Value of Individualised Pacing and Repeated Practice, where the learner felt free to practice with the material multiple times at their own pace. They did this differently than in a normal classroom; listening materials were typically played once or twice for all students simultaneously. "I could listen to the difficult parts many times until I knew how to do it," one participant said. "In a normal class, the teacher plays the audio twice and moves on, but sometimes I need more repetition" (Participant E31, interview). "I liked that I could practice pronunciation every day, not just during class time," she said. This made me accomplish much quicker" (participant E07, interview). Recognition of Authentic Communication Focus was a fourth theme; students reported that intervention activities were about meaningful communication rather than limited skill learning. Some believed that MCOM was an alternative to their past English training: "In school, it was a matter of repeating a sentence from the book. This class taught us how to converse about the subjects of interest to us" (Participant E19, questionnaire). The use of authentic sources was particularly appreciated: "Listening to real TED Talks and news reports rather than just textual dialogue made the learning feel more relevant and motivated me to work harder" (Participant E26, interview). This is the theme that makes sense of what, in the information made available via technology, is meaningful, meaningful work.

This mixed-methods study provides compelling evidence of the impact of MCOM on listening comprehension and speaking skills in online higher education with Arabic-speaking learners. The results of the analyses revealed that the experimental group had a significantly higher level of listening comprehension (mean gain = 12.1 points, d = 1.78) than the control group (4.1 points), indicating that auditory input enhancement via a combination of multimodal supports like transcripts, visual aids, and interactive questions increases comprehension. These gains were sustained at the delayed post-test, suggesting a long-term development in skill rather than short-term performance effects, and extend previous CALL findings to Arabic-speaking learners in digital settings (Mulyadi et al., 2021; Zhou & Wei, 2018). Speaking skills also improved tremendously, particularly in pronunciation and fluency/coherence. Regression analyses indicated that ASR tools like ELSA Speak were the strongest predictors of speaking gains, which validates the phonological development component. This extends previous ASR studies (Sun, 2023; Ngo et al., 2024; Bashori et al., 2024) by situating such tools within a theoretically integrated model, rather than in isolation. A critical analysis of component contributions revealed that the best measure of hearing is using input and the best measure of pronunciation is using speaking gains, supporting the design logic of the MCOM. In particular, interactive oral practice dramatically predicted gains in both areas, indicating the value of authentic, meaning-rich communication amongst holistic oracy constructs consistent with SLA theory (Bailey et al., 2021). Reflective journaling also had a small but important association with speaking development, supporting metacognitive engagement in learning skill acquisition and justifying its inclusion as one of the MCOM components. Qualitative data supported these findings. Students expressed increased confidence and decreased anxiety, which was especially helpful for students from schools with little oral instruction. For example, ASR and AI chatbots provided psychological

DOI: 10.18860 /ijazarabi.V9i2.39744
ISSN(print): 2620-5912 | ISSN(online):2620-5947

safety for experimentation that mirrored the theories of self-determination (Chiu et al. 2024). Multimodal supports, such as audio + transcripts, visual intonation graphs, were especially useful in making abstract English phonological features tangible to compensate for limited L1 transfer.

Participants also appreciated the confluence between practice in the form and meaningful communication, indicating that the content of real-world knowledge (e.g., TED Talks) and meaningful tasks made learning relevant, a common tension in technology-based language instruction. Finally, the implementation challenge focused on practical requirements: free internet, a structured onboarding process to digital tools, and ongoing technical support. Importantly, teachers continued to be present not in the delivery of content but in providing feedback, scaffolding, and personal guidance.

The study has some theoretical implications for the process of oracy acquisition in technology-mediated EFL settings. First, it proves that the combination of Cognitive Load Theory and Multimodal Learning Theory contributes to an effective design framework in addressing the particular issues of the Arabic-speaking learners. Minimising extraneous cognitive load and maximising the use of multiple sensory modalities leads to validated learning results and perceptions about learners. Thus, this form of integration may have a positive impact on future models of such training populations. Second, the study contributes to our knowledge of the roles played by various activities mediated by technology in supporting oracy. The element contribution analysis demonstrates that some components (e.g. auditory input, practice in pronunciation) contribute to target skill gains most significantly. Still, the full oracy development process needs the combined involvement of the elements. This supports the systems perspective of instructional design in which elements do not operate in autonomy, but instead work synergistically. Third, the study contributes to the digital inclusivity field of study by demonstrating that student-centred pedagogies supported by their cultural and linguistic background are highly beneficial.

The MCOM reinforces the pronunciation difficulties of Arabic speakers, its combination of a collectivist (connections with peers) and individualist (self-study) focuses, and its offering of more than one way to prove competence are signs of inclusive design that accommodated a variety of learners. To educators and institutions, there are a number of implications that may be applied practically. First, CALL tools must not be implemented individually but as part of a consistent pedagogical framework. The MCOM is an example of this concept, and it emphasises that technology must be a servant of pedagogy and not its engine. Second, formal, methodical phonological teaching, in particular, the use of ASR tools, is specifically useful in the case of Arabic-speaking students. Even short, daily practice (e.g., 10 minutes), which is maintained for weeks, can lead to significant improvement, which justifies the practice of institutional investment in ASR programs and teacher education on pronunciation instruction. Third, a mix between synchronous and asynchronous activities improves oracy development: asynchronous activities help with repetition, pacing and reflection, and synchronous sessions are the ones that promote the development of authentic interaction and community. These two modes are necessary and are to be deliberately balanced. Fourth, metacognition-based reflection (i.e., guided journaling) is used in terms of assimilating strategies and maintaining independent learning. Such practices may not always be given a priority in time-bound classes, but they should be mentioned. Lasting, practical enablers (e.g., the presence of a solid internet connection, systematic training of orientation to digital tools,

continuing technical assistance, and instructor training) are the keys to successful implementation, however. Institutions should be technologically and humanly prepared to translate sound design into meaningful learning.

Some limitations should be mentioned, and they point to promising future research directions. Quasi-experimental design is contextually a suitable one, limiting causal interference compared to a fully randomised trial. Further investigation, which is randomised and has a larger sample size, will enhance validity. The eight-week intervention showed significant results in the short-term, but would not tell about long-term retention or development to emphasise the necessity of longitudinal studies that would trace gains persistence, compounding, and levelling over time. The results are also limited based on the single-institution environment and relatively homogeneous subject pool. The replication on the different parts of the Arabic-speaking world, education levels, and school contexts would explain the generalizability of the model, and hence, the adjustments. The particular CALL tools employed are also only one implementation of the MCOM framework; comparative studies of appropriate technology configurations within the same theoretical framework may indicate the best tool choices in different situations. Although the emphasis on listening and speaking matches the emphasis on orality, it raises an open question with regard to any possible transfer effects to writing and reading. Future studies could examine the question of whether any of the four skills are indirectly supported by oracy-oriented interventions or could examine model-based approaches that tackle all four skills. Methodologically, the performance tests combined with the questionnaires and interviews might be complemented with multimodal data, including eye-tracking when listening to an activity, discourse analysis of synchronous interactions, or physiological measures of cognitive load to gain a profound understanding of learning processes. Lastly, studies on scaled implementation are required to deal with the real-life issues: instructor training, quality control on multiple sections, cost-efficiency, and sustainability. Although this work establishes efficacy in controlled settings, translating the MCOM into everyday institutional practice is an issue that needs to be addressed by empirical means in terms of feasibility and scale fidelity.

CONCLUSION

This research offers a solid argument that the Multimodal Cognitive Oracy Model is a powerful framework that can be used to improve both listening and speaking skills in Arabic-speaking EFL students in online higher education settings. By blending theoretical concepts of the Cognitive Load Theory and Multimodal Learning Theory with evidence-based CALL practices, a comprehensive approach is formed that will contribute to the special needs of this learner population and promote the inclusive involvement and the availability of skill acquisition. Statistical results indicate significant gains in listening comprehension and speaking proficiency as compared to traditional instruction, and experimental group scores indicated improvements by about 12 points (on 100-point scales), which are effect sizes, and fall in the large range. The qualitative evidence has shown that students feel that the MCOM approach builds confidence, is interesting, and has a relation to their real communication needs. The contribution of the study goes further than proving that a specific intervention is effective, through component contribution analysis and thick description of the experiences of the learners. These observations not only create a theoretical knowledge base of oracy development in technology usage but also a guide to teachers and institutions wishing to improve their

online EFL course offerings. With higher education being digitised, structures such as the MCOM, which carefully unite technology with viable pedagogues and sensitivity regarding learner diversity, will take even greater significance in managing equal access to quality language learning. To Arabic-speaking students who seek to acquire proficiency in English as a gateway to academic and professional achievements, the ability to develop strong oracy competencies is a vital learning requirement. The paper has shown that online instruction that is well-planned can serve this need effectively and help students to acquire the listening and speaking skills they need to engage with people successfully using the English language. When embracing inclusive pedagogies, tapping into appropriate technologies, and remaining focused on meaningful communication and systematic skill building, online EFL education will be able to deliver on its promise of offering accessible, effective language instruction opportunities to different kinds of learners all over the world.

ACKNOWLEDGMENT

The primary researcher would like to thank B.S. Abdur Rahman Crescent Institute of Science and Technology for the stipend and support.

REFERENCES

AbuSahyon, A. S. E., Alzyoud, A., & Alshorman, O. (2023). AI-driven technology and chatbots as tools for enhancing English language learning in the context of second language acquisition: A review study. International Journal of Linguistics, Literature and Translation, 6(11), 234-245. https://doi.org/10.32996/ijllt

Afril, R., Husna, I., & Jasem, H. (2024). Learning Arabic in the digital era: Challenges and effective strategies. International Journal of Arabic Language Education, 1(1), 12-29. https://doi.org/10.1234/ijar.2024.01

Al Fraidan, A., & Alaliwi, M. (2024). Digital transformation for sustainable English language learning: Insights from Saudi Arabia and global perspectives. Forum for Linguistic Studies, 6(2), 1247-1268. https://doi.org/10.30564/fls.v6i2.6234

Au, A., Siddiqi, H., Sayadi, G., Zhao, T., Kleib, M., & Colquhoun, H. (2024). Digital learning preferences of Arabic-speaking older immigrants in Canada: A qualitative case study. Educational Gerontology, 50(8), 673-687. https://doi.org/10.1080/03601277.2024.2370114

Bailey, D., Almusharraf, N., & Hatcher, R. (2021). Finding satisfaction: Intrinsic motivation for synchronous and asynchronous communication in the online language learning context. Education and Information Technologies, 26(3), 2563-2583. https://doi.org/10.1007/s10639-020-10369-z

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2024). 'Look, I can speak correctly': Learning vocabulary and pronunciation through websites equipped with automatic speech recognition technology. Computer Assisted Language Learning, 37(1-2), 222-254. https://doi.org/10.1080/09588221.2022.2080230

Beloufa, C. (2023). How culturally diverse is the virtual space? Towards inclusive pedagogy: A case study at the Arab Open University. In N. Zoghlami et al. (Eds.), Bridging the gaps in language learning (pp. 189-209). Springer. https://doi.org/10.1007/978-3-031-42693-3_10

Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2024). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. Interactive Learning Environments, 32(7), 3240-3256. https://doi.org/10.1080/10494820.2023.2172044

Ghanizadeh, A., Razavi, A., & Jahedizadeh, S. (2015). Technology-enhanced language learning (TELL): A review of resources and upshots. International Letters of Social and Humanistic Sciences, 54, 73-87. https://doi.org/10.18052/www.scipress.com/ILSHS.54.73

Giannakos, M., & Cukurova, M. (2023). The role of learning theory in multimodal learning analytics. British Journal of Educational Technology, 54(5), 1246-1267. https://doi.org/10.1111/bjet.13320

Gilakjani, A. P., Ismail, H. N., & Ahmadi, S. M. (2011). The effect of multimodal learning models on language teaching and learning. Theory and Practice in Language Studies, 1(10), 1321-1327. https://doi.org/10.4304/tpls.1.10.1321-1327

Hopkyns, S. (2023). Toward culturally and linguistically responsive e-learning in post-COVID-19 higher education: Perspectives from the United Arab Emirates. Gulf Education and Social Policy Review, 4(1), 68-93. https://doi.org/10.18502/gespr.v4i1.12608

Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Miao, Y., & Chai, C. S. (2023). Exploring the effects of automatic speech recognition technology on oral accuracy and fluency in a flipped classroom. Journal of Computer Assisted Learning, 39(2), 493-508. https://doi.org/10.1111/jcal.12732

Ma, L. (2025). Investigation of the impact of cognitive load on EFL learners' satisfaction with MOOCs: The mediating role of expectation confirmation and perceived usefulness. BMC Psychology, 13(1), 42-58. https://doi.org/10.1186/s40359-025-02735-8

Malkawi, A. R., Bakar, M. S. A., & Dahlin, Z. M. (2024). Culturally inclusive prototyping for higher education institutions: Navigating language and gender dynamics. Indonesian Journal of Electrical Engineering and Computer Science, 33(3), 1823-1834. https://doi.org/10.11591/ijeecs.v33.i3.pp1823-1834

McCrocklin, S. M. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. System, 57, 25-42. https://doi.org/10.1016/j.system.2015.12.013

Mulyadi, D., Wijayatiningsih, T. D., Singh, C. K. S., & Hoo, G. H. (2021). Effects of technology-enhanced task-based language teaching on learners' listening comprehension and speaking performance. International Journal of Instruction, 14(3), 717-736. https://doi.org/10.29333/iji.2021.14342a

Ngo, T. T. N., Chen, H. H. J., & Lai, K. K. W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. ReCALL, 36(1), 68-88. https://doi.org/10.1017/S0958344023000113

Nguyen, H., Dolan, H., & Taylor, S. (2022). Cultivating intercultural competences in digital higher education through English as an international language. Journal of Academic Language and Learning, 16(2), 48-63. https://journal.aall.org.au/index.php/jall/article/view/771

Ozer, O., & Kılıç, F. (2018). The effect of mobile-assisted language learning environment on EFL students' academic achievement, cognitive load and acceptance of mobile learning tools. EURASIA Journal of Mathematics, Science and

Technology Education, 14(7), 2915-2928. https://doi.org/10.29333/ejmste/90992

Pellicer-Sánchez, A. (2022). Multimodal reading and second language learning. ITL-International Journal of Applied Linguistics, 173(2), 237-261. https://doi.org/10.1075/itl.21039.pel

Satar, H. M. (2015). Sustaining multimodal language learner interactions online. CALICO Journal, 32(3), 480-507. https://doi.org/10.1558/cj.v32i3.26508

Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. Frontiers in Psychology, 14, 1210187. https://doi.org/10.3389/fpsyg.2023.1210187

Takase, N. (2024). Synchronous and asynchronous online communication for developing foreign language speaking skills. Asia CALL Online Journal, 15(1), 45-62. http://asiacall-acoj.org/index.php/journal/article/view/65

Tatlı, Z., Saylan, E., & Kokoç, M. (2022). Digital storytelling in an online EFL course: Influences on speaking, vocabulary, and cognitive load. Participatory Educational Research, 9(3), 446-467. https://doi.org/10.17275/per.22.80.9.3

Yang, S. C., & Chen, Y. J. (2007). Technology-enhanced language learning: A case study. Computers in Human Behavior, 23(1), 860-879. https://doi.org/10.1016/j.chb.2006.02.015

Yu, Z., Ying, Z., & Zhichun, Y. (2019). Student satisfaction, learning outcomes, and cognitive loads with a mobile learning platform. Computer Assisted Language Learning, 32(4), 323-341. https://doi.org/10.1080/09588221.2018.1517093

Zhou, Y., & Wei, M. (2018). Strategies in technology-enhanced language learning. Studies in Second Language Learning and Teaching, 8(2), 471-495. https://doi.org/10.14746/ssllt.2018.8.2.13