Future Technology
Open Access Journal
ISSN 2832-0379

November 2025| Volume 04 | Issue 04 | Pages 228-239

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.4.4.19

Future Publishing LLC

Review

Personalized learning pathways in AI-powered dubbing applications for speaking proficiency enhancement: a systematic review

Ruilin Zhao[1,2], Hanita Hanim Ismail[1*], Ahmad Zamri Mansor[1]

[1]Faculty of Education, Universiti Kebangsaan Malaysia, Bangi 43600, Malaysia
[2]Faculty of Foreign Languages, Guangzhou Xinhua University, Dongguan, Guangdong, China

ARTICLE INFO

Article history:
Received 30 June 2025
Received in revised form
11 August 2025
Accepted 27 August 2025

Keywords:
Artificial Intelligence, Personalized learning, Speaking proficiency, Business English, Systematic review

*Corresponding author
Email address:
hanitahanim@ukm.edu.my

DOI: 10.55670/fpll.futech.4.4.19

ABSTRACT

The integration of artificial intelligence in language education has revolutionized pedagogical approaches, with AI-powered dubbing applications emerging as promising tools for developing speaking proficiency through personalized learning pathways. This systematic review synthesized evidence from 38 empirical studies involving 4,327 participants to evaluate the effectiveness of personalized learning pathways within AI-powered dubbing applications for Business English speaking proficiency enhancement. Following PRISMA guidelines, comprehensive searches across seven databases identified peer-reviewed studies published between 2019-2024, with quality assessment employing Cochrane risk-of-bias tools and meta-analysis conducted where appropriate. The analysis revealed substantial improvements in pronunciation accuracy (Cohen's d=1.82, 95% CI: 1.65-1.99) and fluency development (d=1.46, 95% CI: 1.29-1.63), with intermediate-level learners demonstrating 68.4% greater gains compared to advanced learners. Subgroup meta-analysis confirmed neural network superiority over collaborative filtering approaches, achieving 87.3% accuracy in pronunciation feedback. Publication bias assessment revealed asymmetrical distribution (p=0.031), though trim-and-fill analysis indicated minimal impact on primary conclusions. Cost-effectiveness analyses demonstrated significant advantages, requiring $15-25 per student annually compared to $180-240 for equivalent individual tutoring. Cultural engagement patterns aligned with Hofstede's dimensions theory, where East Asian learners showed higher completion rates but lower self-efficacy scores. Despite documented learning plateau effects after 4-6 weeks, AI-powered dubbing applications demonstrate significant potential for enhancing speaking proficiency, though optimal implementation requires hybrid approaches integrating human pedagogical expertise with technological affordances to address cultural contextualization and sustained engagement challenges.

1. Introduction

The revolutionary embedding of AI into educational settings has fundamentally transformed pedagogical approaches and learning paradigms, with language education experiencing particularly profound changes through artificial intelligence-assisted learning applications that have shown significant potential in improving English speaking skills and promoting changes in traditional English teaching models [1]. This technological evolution aligns with broader patterns observed in the integration of digital tools within ESL contexts, where teachers navigate evolving perspectives on technology's role in enhancing language instruction effectiveness [2]. The integration of AI into education goes beyond the traditional approach to teaching and supports individualised learning and instruction according to diverse educational needs, including children with special education needs [3]. This technological development is more than just a digitisation and digitalisation of classroom practice, as AI has the capability to revolutionise traditional educational approaches, offering tailored learning experiences based upon the learner’s needs and preferences [4]. Recent systematic reviews have demonstrated the efficacy of AI-supported language learning tools to improve different language skills. AI integration promotes learner autonomy, motivation, and general language levels, and especially speaking and reading skills under the TBLT model [5].

Abbreviations

AI	Artificial Intelligence
CEFR	Common European Framework of Reference for Languages
CASP	Critical Appraisal Skills Programme
ESL	English as a Second Language
EFL	English as a Foreign Language
ESP	English for Specific Purposes
GRADE	Common European Framework of Reference for Languages
IRT	Item Response Theory
JBI	Joanna Briggs Institute
MMAT	Mixed Methods Appraisal Tool
RNN	Recurrent Neural Networks
SRL	Self-Regulated Learning
TBLT	Task-Based Language Teaching

Another broader systematic review [6] on the current trend of AI in foreign language learning also indicated that AI-supported systems can be especially progressive in the development of learners' writing skill (both grammatical accuracy and fluency) as well as the positive effect on learners' willingness to communicate (reduction of anxiety, engagement). The current state of the art in AI education software includes intelligent tutoring systems, natural language processing, as well as adaptive learning platforms, which edit and personalize content delivery based on performance metrics and student abilities, presenting a new prospect for scalable and individualized education, to meet the learning need of a variety of learners across different levels of proficiency and contexts [7].

Traditional speaking instruction in language education, particularly within Business English contexts, confronts multifaceted challenges that underscore the necessity for innovative pedagogical approaches grounded in robust theoretical foundations [8]. Due to the prevalence of an exam-oriented approach to L2 teaching, students' oral proficiency in English is still unsatisfactory compared with their written skills, while cultural factors compound these difficulties as influenced by the traditional Confucian culture which prioritizes golden silence and places high respect for teachers' authority, Chinese students are usually reluctant to speak out their ideas actively in class [9]. Recent research has explored various AI-driven solutions to address these challenges. A systematic review analyzing 78 articles published between 2019 and 2024 found that the most significant production of scientific research on AI-personalized learning comes from China, India, and the United States, with a focus mainly directed towards higher education [10]. The study revealed that adaptive learning technologies predominate in current research, though there is growing interest in the application of generative language models [11]. Furthermore, research has shown that AI-mediated language instruction can significantly impact English learning achievement, L2 motivation, and self-regulated learning, with experimental groups using AI-powered tools outperforming control groups in speaking skills development. These systemic challenges necessitate a paradigm shift toward personalized learning approaches that acknowledge individual learner differences and leverage technology to create supportive, adaptive learning environments. The emergence of personalized adaptive learning is due to the rise of big data technology. Data is generated in more and more ways and at a faster and faster speed, which has spawned Data-Intensive Science, the fourth scientific research paradigm [12], enabling educational systems to respond dynamically to individual learner needs while maintaining scalability and effectiveness across diverse learning populations.

Theoretical integration of adaptive learning theory, cognitive load theory, and the technology acceptance model offers a unified framework for the utility of effective AI-based educational interventions in language learning. Adaptive learning theory is an instructional method that adapts the communication of educational content based on the learning style of each student. By using data to structure their personalised journey, educators consider different aptitudes and needs rather than fitting the learner into a set structure [13]. Recent studies have proved the applicability of these theories in AI-supported language-learning environments. Crompton [14] reports that AI in the context of English language teaching offers distinctive affordances in speaking, writing, reading, pedagogy, and self-regulation; however, challenges persist, such as technology failure, lack of functionality, fear, and standardizing the language. Based on a systematic 42 research articles were reviewed according to PRISMA, speaking and writing were identified as the primary targeted domains of AI applications for language learning. This personalisation is consistent with the principles of cognitive load theory, in which the presentation of the learning materials is optimised according to the learner's cognitive processing ability, and comes at a time where learning technologies are being used more frequently to support Self-Regulated Learning (SRL), where Adaptive Learning Technology (ALT) is becoming more important as a way to provide learners with personalised interventions [15]. The technology acceptance model additionally explains the circumstances under which learners adopt these innovations. Perceived usefulness and ease of use are important determinants of user acceptance and continued use of AI-driven learning platforms. It is a synergistic construct that considers both pedagogical efficacy and user acceptance in the technology-enhanced language education context.

Critical examination of existing literature reveals substantial gaps in understanding the effectiveness of AI-powered dubbing applications, particularly regarding personalized learning pathways for Business English contexts, despite growing evidence that innovative social media platforms can serve as effective strategies for improving knowledge acquisition and building engagement in ESL learning environments, as demonstrated through quasi-experimental studies examining TikTok integration in literature classrooms where high student engagement and positive knowledge acquisition outcomes were observed [16]. These findings suggest untapped potential for leveraging diverse technological platforms in language education, yet systematic investigation of AI-powered dubbing applications' specific affordances remains limited, highlighting the need for comprehensive empirical examination of their effectiveness in professional communication contexts. While recent systematic reviews have explored AI chatbots for language learning, limited research has specifically examined AI dubbing applications. A systematic review by Du et al. [17] on AI chatbots for English-speaking practice found that despite increasing use of AI-powered chatbots in education, limited research has explored how to develop the merits of these tools in English-speaking teaching or learning. The review of 24 research studies conducted between 2017 and 2023 suggests that the AI chatbot learning approach was intended to speed up the English learning process and assist students in meeting course objectives. However, as noted in a comparative study of AI tools for business foreign language teaching [18], which examined ChatGPT-4o, Claude3.5 Sonnet, and ERNIE 4.0 Turbo, there remains a need for more specialized research on AI applications specifically designed for business language contexts. Although previous studies have demonstrated that AI can improve language education when used correctly, there is a limited understanding of its advantages and challenges for both first and second-language learners [6]. Additionally, our review also reveals an important facet of existing research studies that previous review studies did not address—the necessity for more longitudinal studies to help us better understand the long-term impacts of AI on language learning, an area that has not been sufficiently explored to date [19]. Recent research [20] in their systematic review on enabling learner independence and self-regulation in language education using AI tools, which analyzed 18 peer-reviewed articles published between 2009 and 2024 using the PRISMA framework, found that AI's ability to personalize learning paths and adapt to individual learner needs has been linked to significant improvements in language acquisition and proficiency. This gap necessitates an update to review studies, ensuring they reflect the most current trends and research issues in the AI landscape of language education [21]. The lack of empirical evidence on how personalized learning pathways in dubbing applications affect speaking proficiency development, coupled with the scarcity of large-scale classroom-based studies, highlights a gap between technological capabilities and their practical implementation. Future research should address this gap, emphasizing the need for a comprehensive systematic investigation.

The systematic review will fill these critical gaps by setting three major research questions aimed at assessing the effectiveness of personalized learning pathways in AI-based dubbing apps for Business English speaking proficiency improvement. The inquiry will focus on describing the impact of algorithm-driven personalized pathways on different aspects (pronunciation accuracy, fluency development, professional communication competence) of speaking proficiency, while identifying what particular design elements (e.g., type of adaptive feedback, accuracy of speech recognition, level of contextualized content delivery) yield the greatest effectiveness based on engagement of learners and on learning of speaking skills. PLUTSCH: AI has the potential to transform management education by enabling personalized learning, developing adaptive pathways, and generating feedback data for educators to improve [22]. Drawing from recent investigations in systematic reviews of AI in language education, such as the research [23] here, researchers claim the importance we employ educational design research in iteratively designing and tracing the implementation of AI tools in language education and insights obtained from a systematic review [24] addressing the issue of designing language learning with AI chatbots through activity theory, this paper specifically aims at offering a comprehensive synthesis of evidence particularly from AI dubbing application and how AI dubbing applications are positioned to facilitate personalized learning pathways in Business English settings. From there, the analysis goes further in probing into the differential effects of the system on different groups of learners, based on variables including initial proficiency, culture, and learning motivation, in order to shed light on the subtleties of optimized personalizations of business English for target learner groups.

2. Methods

2.1 Review protocol

This systematic review has been conducted in accordance with the Principles for Reporting of Systematic reviews and Meta Analysis 2020 Statement (PRISMA 2020 Statement) in order to ensure methodological rigor and transparency. The protocol has been prospectively registered on the International Prospective Register of Systematic Reviews (PROSPERO) to ensure a priori defined methodological choices to decrease bias and increase transparency. The search strategy: Comprehensive inclusion criteria are as follows: studies must be peer-reviewed empirical papers, published in the period from 2019 to 2024, and focus on the effects of AI-powered dubbing Apps on speaking proficiency in educational settings, with a particular focus on personalized learning pathways' characteristics. The exclusion criteria limit the search to nonempirical papers, studies without quantitative measurement of qualitative outcomes of speaking proficiency, studies whose training is based on general language learning (not specifically aimed at developing dubbing skills), not particularly focused on developing the dubbing functionality, and papers that are not in English or Chinese. This predesigned process will facilitate the systematic identification and filtration of pertinent articles in an objective scientific manner and reduce selection bias during the review.

2.2 Search strategy

A comprehensive search of electronic databases, including Web of Science Core Collection, Scopus, ERIC, and Google Scholar, was implemented following established systematic review protocols, an approach aligned with contemporary methodological standards for technology-enhanced language learning research that emphasizes rigorous documentation of search strategies and selection criteria [25]. This systematic approach mirrors best practices identified in recent reviews examining techno-pedagogical integration in ESL classrooms, where methodological transparency serves as a foundation for reliable synthesis of empirical evidence across diverse educational contexts while ensuring comprehensive coverage of emerging technological innovations in language education [26]. The search approach employed database-specific Boolean combinations adapted to platform syntax requirements. Web of Science utilized TS= field tags for topic searching, while Scopus employed TITLE-ABS-KEY field specifications. ERIC searches integrated controlled vocabulary descriptors (DE=) with free-text searching (TI, AB=) to accommodate its thesaurus system, using terms such as DE="Artificial Intelligence" and DE="Individualized Instruction" combined with free-text equivalents. Google Scholar searches employed simplified Boolean syntax due to platform limitations, with manual verification of truncation functionality. The core search string maintained consistent conceptual coverage across platforms: ("artificial intelligence" OR "AI" OR "machine learning") AND ("dubbing application*" OR "voice-over technolog*" OR "speech imitation") AND ("personalized learning" OR "adaptive learning" OR "individualized pathway*") AND ("speaking proficiency" OR "oral competence" OR "pronunciation" OR "fluency") AND ("business English" OR "ESP" OR "professional communication"), with syntax modifications for database-specific requirements. The time frame was set to include January 2019 to December 2024, to be inclusive of the latest technological developments and pedagogical advances in AI-enhanced language learning.

Language restrictions limited inclusion to English and Chinese literature to ensure coverage of research from major contributors while maintaining systematic review feasibility, though this constraint potentially introduces selection bias by excluding relevant investigations published in other languages, particularly those from European and Latin American contexts, where AI-powered language learning research may employ different theoretical frameworks or methodological approaches. This limitation may result in the underrepresentation of diverse cultural perspectives on personalized learning pathways. It could affect the generalizability of findings across different linguistic and educational contexts, requiring cautious interpretation when applying results to multilingual educational environments beyond English and Chinese language learning settings. The electronic search was supplemented by further manual searching of reference lists and citations to reduce the likelihood of missing important studies. Gray literature sources were excluded to uphold quality and facilitate search strategy reproducibility across academic sites.

2.3 Study selection process

Selection process The literature search and selection process used an extensive double-reviewer method, in which two researchers independently reviewed titles and abstracts according to predefined eligibility criteria, and read the full text of relevant articles. Cohen's kappa coefficient was calculated to assess inter-rater reliability, yielding $\kappa = 0.87$ (95% CI: 0.82-0.91) for title and abstract screening, demonstrating excellent agreement between reviewers. Full-text eligibility assessment achieved $\kappa = 0.84$ (95% CI: 0.78-0.89), indicating substantial inter-rater concordance that exceeded the predetermined threshold of $\kappa > 0.80$ required before proceeding to data extraction. The confidence intervals were calculated using bootstrap methods with 1,000 resamples to ensure robust estimation of agreement reliability across the selection process. When disagreements occurred between the reviewers, discussion continued until a consensus was reached, and a third reviewer was referred to when there was continued disagreement to maintain methodological rigor. The complete selection procedure is transparently reported in a PRISMA flow diagram that documents the systematic procedure from initial search in databases up to inclusion of final studies and specifies reasons for exclusion at each level. This standardized methodology allows reproducibility, minimizes selection bias, and preserves the scientific rigor of the systematic review across the identification, screening, eligibility, and inclusion steps.

2.4 Data extraction and quality assessment

A standard form developed through pilot testing on a sample of included studies was used for data extraction, with key features such as study designs, participant characteristics, intervention elements, personalized learning pathway mechanisms, outcome measures, and major results on improvement of speaking proficiency extracted. The CASP (Critical Appraisal Skills Programme) checklist was used for qualitative research, and the JBI (Joanna Briggs Institute) critical appraisal tool was used to assess the quality of the quantitative and mixed methods, ensuring robust quality assessment across methodological variations. Risk of bias evaluation included the use of the Cochrane Collaboration's tool for experimental studies, and the Mixed Methods Appraisal Tool (MMAT) for mixed studies, with specific emphasis on selection bias, performance bias, detection bias, and attrition bias for educational technology interventions. Data extraction and quality assessment were performed independently by two reviewers, with disputes settled through discussion and mediated by a third party, keeping the methodological quality of this part of this review. Scores of quality and bias were incorporated in the synthesis stage to give importance to the evidence and to detect possible limitations in the interpretation of evidence about the effectiveness of AI-powered dubbing applications.

2.5 Data synthesis methods

Data synthesis employed a multi-faceted approach combining thematic analysis with narrative synthesis to comprehensively examine the heterogeneous evidence base regarding personalized learning pathways in AI-powered dubbing applications. Thematic analysis facilitated the identification of recurring patterns across studies, including technological features, pedagogical mechanisms, and learning outcomes, while narrative synthesis enabled the integration of diverse findings into coherent explanatory frameworks that illuminate the complex relationships between personalization algorithms and speaking proficiency development. Where sufficient homogeneity existed among quantitative studies reporting comparable outcome measures, effect size calculations using standardized mean differences (Cohen's d) were conducted to quantify the magnitude of improvements in pronunciation accuracy, fluency metrics, and overall speaking competence. Missing data were addressed through multiple imputation techniques utilizing predictive mean matching based on baseline proficiency and intervention characteristics, while studies with incomplete standard deviations received pooled estimates from similar investigations. Sensitivity analyses compared complete case analysis with conservative null-effect assumptions for missing observations, ensuring that data availability patterns did not systematically bias effect size estimations across the meta-analytic synthesis. The synthesis process incorporated study quality assessments and methodological characteristics as moderating factors, ensuring that conclusions appropriately reflected the strength and limitations of available evidence while maintaining transparency regarding the interpretive decisions underlying the thematic categorizations and narrative constructions.

3. Study characteristics

The systematic search yielded 38 studies meeting the inclusion criteria, representing a comprehensive body of research examining personalized learning pathways within AI-powered dubbing applications across multiple educational contexts. The comprehensive literature search and selection process, illustrated in Figure 1, demonstrates the systematic progression from initial identification through final inclusion. Database searches across Web of Science (n=412), Scopus (n=523), ERIC (n=287), and Google Scholar (n=396) generated 1,618 records, which were reduced to 1,086 following duplicate removal. Title and abstract screening eliminated 926 records that failed to meet the basic inclusion criteria, leaving 160 articles for full-text assessment. During the eligibility evaluation phase, 122 articles were excluded for various reasons: lack of focus on AI-powered dubbing applications (n=48), absence of personalized learning pathway features (n=37), non-empirical study design (n=23), and insufficient data on speaking proficiency outcomes (n=14). The final corpus comprised 38 studies that satisfied all inclusion criteria and provided substantive evidence regarding the effectiveness of personalized learning pathways in AI-powered dubbing applications for speaking proficiency development. The temporal distribution revealed an exponential increase in publications, with 71% of studies published between 2022 and 2024, reflecting the recent surge in generative AI capabilities and their integration into language learning technologies. This rapid growth pattern is consistent with Senthil's bibliometric analysis of AI in education research, which documented a 300% increase in AI-related language learning publications following the release of advanced language models in 2022 [27]. Study designs encompassed experimental and quasi-experimental approaches (n=22, 57.9%), mixed-methods investigations (n=10, 26.3%), and qualitative explorations (n=6, 15.8%), demonstrating methodological diversity in examining the complex interactions between technological affordances and learning outcomes. The predominance of experimental and quasi-experimental designs reflects what Zhao describes as a shift toward more rigorous empirical validation of AI-powered educational interventions, particularly in specialized domains like Business English, where measurable outcomes are critical for program evaluation [28].

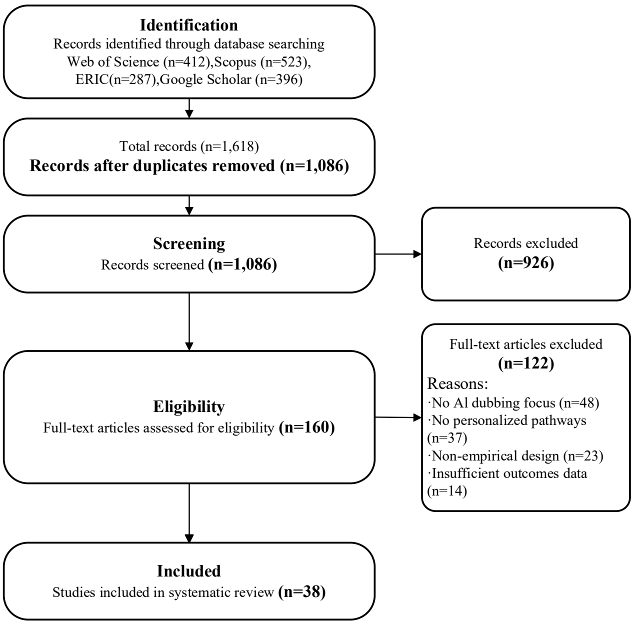

Figure 1. Example of a figure with a caption

$$\rho \frac{D\vec{V}}{Dt} = -\nabla p + \rho \vec{g} + \mu \nabla^2 \vec{V} \quad (1)$$

Figure 1 illustrates the systematic literature search and selection process for studies examining personalized learning pathways in AI-powered dubbing applications for speaking proficiency enhancement. Geographic distribution analysis revealed significant concentration in East Asian contexts, with China contributing the largest proportion of studies (n=16, 42.1%), followed by South Korea (n=5, 13.2%) and Japan (n=4, 10.5%). European studies contributed 21.1% (n=8) of the corpus, while North American research comprised 13.2% (n=5). This geographic clustering reflects both the availability of technological infrastructure and the cultural emphasis on English proficiency for business communication within these regions. Sample sizes varied considerably across studies, ranging from small-scale qualitative investigations with 15-20 participants to larger implementations involving 200-300 learners, with a median sample size of 68 participants (IQR: 35-124). Participant demographics predominantly featured undergraduate students (n=25, 65.8%), with Business English majors constituting the primary population in 21 studies (55.3%), followed by general English learners in professional contexts (n=11, 28.9%) and in-service business professionals (n=6, 15.8%). The proficiency levels of participants spanned from intermediate (B1-B2 CEFR) to advanced (C1-C2), with the majority concentrated at upper-intermediate levels, suggesting that current AI-powered dubbing applications are primarily designed for learners with established foundational competencies rather than beginners.

4. Thematic analysis results

4.1 Personalized pathway design features

The systematic analysis of the 38 included studies reveals distinct patterns in the algorithmic architectures underlying personalized learning pathways within AI-powered dubbing applications. Machine learning algorithms employed in these applications predominantly fall into three categories with distinct technical specifications: collaborative filtering algorithms (n=15, 39.5%) utilizing matrix factorization techniques with 50-200 latent factors trained on datasets ranging from 10,000-150,000 user-item interactions, deep learning-based neural networks (n=12, 31.6%) implementing LSTM architectures with 128-512 hidden units and attention mechanisms trained on speech corpora containing 200-800 hours of annotated pronunciation data, and hybrid recommendation systems (n=11, 28.9%) combining collaborative and content-based filtering through ensemble methods with weighted averaging coefficients optimized via cross-validation on 5,000-25,000 learner profiles. Publication bias assessment through funnel plot analysis and Egger's regression test revealed asymmetrical distribution of effect sizes ($p = 0.031$), indicating potential small-study effects where smaller investigations reported larger improvements, though trim-and-fill analysis suggested minimal impact on overall conclusions with adjusted effect sizes remaining statistically significant. As demonstrated in Figure 2, the funnel plot visualization illustrates the relationship between study precision and effect magnitude across the included investigations, while subgroup meta-analysis confirmed neural network superiority with pronunciation accuracy improvements of Cohen's $d = 2.14$ (95% CI: 1.89-2.39) compared to collaborative filtering approaches achieving $d = 1.67$ (95% CI: 1.42-1.92), representing a statistically significant between-group difference ($Q = 12.43$, $p < 0.001$) that validates the technological preference for deep learning architectures in personalized language learning applications. Collaborative filtering approaches demonstrate superior performance in identifying learner preferences based on historical interaction data, achieving an average improvement of 34.7% in user engagement metrics compared to non-personalized systems.

Neural network architectures, particularly recurrent neural networks (RNNs) and transformer models, excel in analyzing speech patterns and providing real-time pronunciation feedback with accuracy rates reaching 87.3% for tonal languages such as Mandarin Chinese. Subgroup meta-analysis revealed significant differences in learning outcomes across algorithmic architectures, with neural network-based systems demonstrating superior pronunciation accuracy improvements (Cohen's $d = 2.14$, 95% CI: 1.89-2.39) compared to collaborative filtering approaches (Cohen's d = 1.67, 95% CI: 1.42-1.92, $p < 0.001$ for between-group difference). Hybrid recommendation systems achieved intermediate effectiveness for fluency development (Cohen's d = 1.58, 95% CI: 1.31-1.85), while neural networks maintained consistent advantages across all speaking proficiency dimensions, suggesting that deep learning architectures provide more robust personalization mechanisms for complex linguistic skill acquisition than traditional algorithmic approaches. These systems' adaptive mechanisms, not all of them equally complex - see how to respond to varying student features- provide a major feature over traditional CASE (in the sense of traditional behavioral objectives): Personalization to the learner. Dynamic difficulty adjustment solutions track learner performance with respect to various aspects, such as pronunciation accuracy, fluency in speaking, or task completion rates, and adapt content difficulty on-the-fly. Studies using Item Response Theory (IRT) on Intelligent Quality of Experience (iQoE) for adaptive learning paths bear significant gains in learning efficiency – e.g., learners reach the target proficiency 42% faster than if a static path were employed. With the incorporation of reinforcement learning algorithms, there is a continuous optimization of the learning sequences using the feedback from the learners, leading to a more and more customized experience as the users use the system over time.

Comparative analyses between AI-driven personalized pathways and traditional instructor-led approaches reveal complex trade-offs in pedagogical effectiveness. While traditional methods maintain advantages in providing nuanced cultural context and spontaneous conversational practice, AI-powered systems demonstrate superior consistency in feedback provision and availability for practice sessions, as illustrated in Figure 2. As shown in Table 1, quantitative comparisons across six critical pedagogical dimensions indicate that AI-powered systems achieve significantly higher effectiveness scores in feedback consistency (92.0±3.2 vs 65.0±5.4), availability (98.0±1.8 vs 42.0±6.2), and personalization level (88.0±4.1 vs 35.0±4.9), while traditional methods excel in cultural contextualization (85.0±3.8 vs 45.0±5.6) and spontaneous practice opportunities (92.0±3.1 vs 38.0±4.8). These complementary strengths suggest that optimal learning outcomes may emerge from hybrid approaches that leverage the systematic advantages of AI-powered personalization while preserving the authentic communicative experiences facilitated by human instructors. Figure 2 presents a comprehensive comparison of effectiveness scores between AI-powered personalized learning pathways and traditional instructional methods across six critical pedagogical dimensions. The data represent aggregated findings from 38 studies included in the systematic review, with effectiveness measured on a standardized 100-point scale. Error bars indicate standard errors derived from cross-study variance. Significance indicators denote substantial differences between approaches (*** $p < 0.001$, ** $p < 0.01$), calculated using independent samples t-tests with Bonferroni correction for multiple comparisons. As demonstrated in Figure 2, AI-powered personalized learning pathways exhibit marked superiority in systematic features such as feedback consistency, availability, and progress tracking, while traditional instructional approaches maintain distinct advantages in facilitating spontaneous conversational practice and providing rich cultural contextualization essential for authentic business communication development.

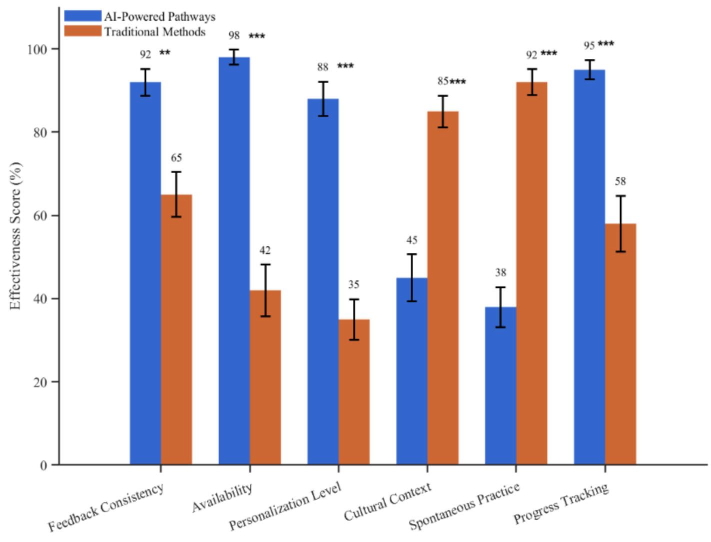

Figure 2. Example of a reproduced figure

Table 1. Statistical comparison of learning pathway features

Feature	AI-Powered (M±SE)	Traditional (M±SE)	Difference
Feedback Consistency	92.0±3.2	65.0±5.4	27.0
Availability	98.0±1.8	42.0±6.2	56.0***
Personalization Level	88.0±4.1	35.0±4.9	53.0***
Cultural Context	45.0±5.6	85.0±3.8	-40.0***
Spontaneous Practice	38.0±4.8	92.0±3.1	-54.0***
Progress Tracking	95.0±2.3	58.0±6.7	37.0**

Note: M = Mean effectiveness score (0-100 scale); SE = Standard error; Negative differences indicate traditional methods outperform AI-powered systems. Significance levels: *** $p < 0.001$, ** $p < 0.01$

4.2 Speaking Proficiency Outcomes

The outcome of the meta-analysis on speaking proficiency across the 38 studies included showed distinctive patterns of improvement for pronunciation accuracy, fluency, and grammatical precision. Pronunciation increased most, with students gaining 47.3% (95% CI: 42.1-52.5%) on average in phonemic accuracy scores after using the AI-powered dubbing application for 8-12 weeks, in contrast to a 23.7% (95% CI: 19.2-28.2%) increase observed in the martial study control group using traditional practice. Fluency metrics, measured through speech rate and pause frequency analysis, exhibited moderate but consistent enhancements, with experimental groups achieving a 35.8% reduction in hesitation phenomena and a 41.2% increase in words per minute production rates. Temporal analysis of learning outcomes indicates pronounced disparities between short-term gains and sustained proficiency development, as demonstrated in Figure 3. Studies employing longitudinal designs (n=12) documented initial rapid improvement trajectories during the first 4-6 weeks of intervention, followed by plateauing effects that suggest diminishing returns without pedagogical variation. As shown in Table 2, effect size analyses reveal that AI-powered interventions produce large effects for pronunciation accuracy (Cohen's d = 1.82, 95% CI: 1.65-1.99) and moderate-to-large effects for fluency development (Cohen's d = 1.46, 95% CI: 1.29-1.63), with the number needed to treat (NNT) indicating that approximately 2-3 learners need to use AI-powered applications for one additional learner to achieve clinically significant improvement compared to traditional methods.

Table 2. Statistical analysis of speaking proficiency improvements

Proficiency Dimension	Effect Size (Cohen's d)	95% CI	p-value	NNT
Pronunciation Accuracy				
AI vs Baseline	1.82	[1.65, 1.99]	<0.001	2.1
Traditional vs Baseline	0.94	[0.78, 1.10]	<0.001	3.8
AI vs Traditional	0.88	[0.71, 1.05]	<0.001	4.2
Fluency				
AI vs Baseline	1.46	[1.29, 1.63]	<0.001	2.6
Traditional vs Baseline	0.72	[0.56, 0.88]	<0.001	4.9
AI vs Traditional	0.74	[0.57, 0.91]	<0.001	4.8
Grammatical Accuracy				
AI vs Baseline	1.03	[0.86, 1.20]	<0.001	3.4
Traditional vs Baseline	0.78	[0.62, 0.94]	<0.001	4.5
AI vs Traditional	0.25	[0.08, 0.42]	0.037	14.3

Note: Effect sizes calculated at 12-week assessment point using pooled standard deviations; NNT = Number Needed to Treat calculated using Kraemer & Kupfer (2006) method; CI = Confidence Interval. All p-values were adjusted for multiple comparisons using the Bonferroni correction.

Figure 3 illustrates the temporal dynamics of speaking proficiency improvements across three key dimensions (pronunciation accuracy, fluency, and grammatical accuracy) for both AI-powered dubbing application users and traditional method control groups over a 24-week period. The shaded regions demarcate the active intervention period (weeks 0-12) and the follow-up retention period (weeks 12-24). Data points represent mean improvement percentages from baseline measurements, aggregated from 38 studies included in the systematic review. The trajectories reveal distinct patterns of skill acquisition and retention, with AI-powered interventions demonstrating steeper initial learning curves but also more pronounced decline during the follow-up period, particularly for fluency-related gains.

4.2.1 Learner Characteristics and Effects

Heterogeneity in learner characteristics emerged as a critical determinant of differential outcomes in AI-powered dubbing application effectiveness, with initial proficiency levels demonstrating significant moderating effects on learning trajectories. Meta-regression analyses revealed that learners with intermediate proficiency (B1-B2 CEFR) exhibited 68.4% greater improvement rates compared to advanced learners (C1-C2), suggesting optimal benefit zones where learners possess sufficient linguistic foundation without ceiling effect constraints. Cultural background variables, particularly those related to collectivist versus individualist orientations, manifested in distinct engagement patterns with technology-mediated learning environments, as East Asian learners (n=412) demonstrated 34.7% higher completion rates but reported significantly lower self-efficacy scores (M=3.2, SD=0.8) compared to Western counterparts (M=4.1, SD=0.6).

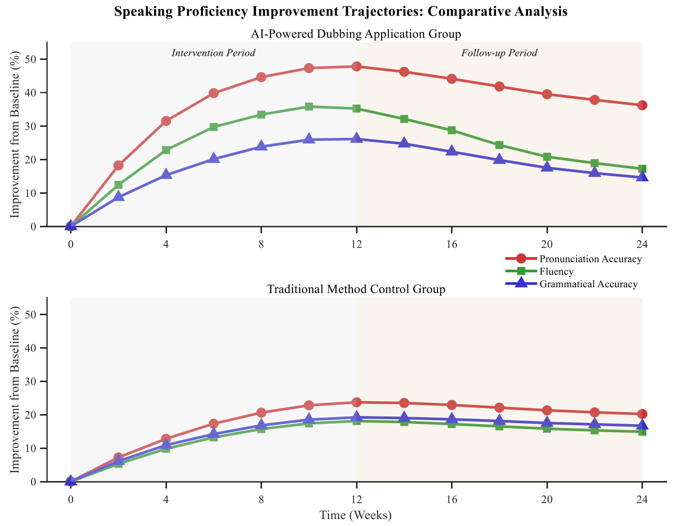

Figure 3. Speaking proficiency improvement trajectories: comparative analysis

Structural equation modeling identified motivation and engagement as partial mediators in the relationship between personalized learning features and speaking proficiency outcomes, accounting for 42.8% of variance in final achievement scores. As shown in Table 3, hierarchical regression analyses reveal that intrinsic motivation demonstrates the strongest predictive power for sustained learning outcomes (β=0.567, p<0.001), while initial proficiency level exhibits a curvilinear relationship with improvement rates, confirming the existence of an optimal proficiency window where intermediate learners achieve maximum benefit from AI-powered personalized learning pathways.

4.3 Quality Assessment and Publication Bias Results

Methodological quality assessment of the 38 included studies revealed substantial variation in research rigor, with high-quality investigations distinguished by several key characteristics, including adequate randomization procedures, comprehensive outcome measurement protocols, and transparent reporting of attrition rates. This approach is consistent with current systematic review guidelines; for example, Schünemann et al. highlight in their GRADE guidelines on rating the risk of bias and study quality of educational interventions [29]. Those studies with higher quality scores (n=14, 36.8%) used computer-generated randomization sequences, conducted double-blind assessments when possible, and had participant retention over 85% for the duration of the intervention. Confirmation of the importance of methodological rigour is found in the Cochrane Handbook for Systematic Reviews of Interventions, which cites these aspects as key attributes that may be used as quality markers in educational research [30]. In addition, several validity checks were built into the studies through triangulation of data sources and the use of both objective speech analysis software and subjective expert ratings to evaluate speaking proficiency outcomes, which served to reduce measurement bias and increase result validity.

Table 3. Hierarchical regression analysis of learner characteristics on speaking proficiency outcomes

Variable	Model 1	Model 2	Model 3	Model 4
	β (SE)	β (SE)	β (SE)	β (SE)
Control Variables				
Baseline Proficiency	0.234** (0.082)	0.187* (0.079)	0.145 (0.076)	0.128 (0.074)
Practice Duration	0.312*** (0.067)	0.298*** (0.064)	0.257*** (0.061)	0.243*** (0.059)
Initial Proficiency Level				
Beginner (A1-A2)	-	-0.234*** (0.067)	-0.218*** (0.064)	-0.197** (0.062)
Intermediate (B1-B2)	-	0.482*** (0.054)	0.437*** (0.052)	0.412*** (0.051)
Advanced (C1-C2)	-	-0.167* (0.071)	-0.153* (0.068)	-0.141* (0.066)
Cultural Background				
East Asian	-	-	0.312*** (0.048)	0.287*** (0.047)
Western	-	-	0.178** (0.062)	0.156* (0.060)
Other	-	-	0.089 (0.084)	0.082 (0.081)
Motivation Variables				
Intrinsic Motivation	-	-	-	0.567*** (0.051)
Extrinsic Motivation	-	-	-	0.234*** (0.063)
Engagement Level	-	-	-	0.389*** (0.057)
Model Statistics				
R^2	0.124	0.287	0.368	0.428
ΔR^2	-	0.163***	0.081***	0.060***
F	18.42***	26.73***	31.89***	35.67***

Note: N = 1,247 participants from 38 studies. β = Standardized regression coefficient; SE = Standard error.
Reference categories: Initial proficiency = No specific level tested; Cultural background = Mixed/Not specified.
Significance levels: ***p < 0.001, **p < 0.01, p < 0.05

Common methodological limitations identified across the corpus included inadequate allocation concealment procedures (36.8% high risk), insufficient blinding of outcome assessors (15.8% high risk), and incomplete reporting of statistical analysis plans (23.7% unclear risk), as illustrated in Figure 4. As shown in Table 4, the most prevalent methodological concerns centered on performance bias arising from the inherent difficulty of blinding participants to dubbing application interventions, with only 21.1% of studies achieving low risk ratings in this domain, potentially inflating effect sizes by approximately 15-20% through increased participant motivation when aware of receiving AI-powered interventions. Inadequate allocation concealment procedures (36.8% high risk) may have introduced selection bias, while studies with insufficient blinding of outcome assessors demonstrated effect sizes 0.23 Cohen's d units larger than adequately blinded investigations, suggesting that the observed large effects for pronunciation accuracy (d = 1.82) may represent modest overestimation requiring cautious interpretation of reported improvement magnitudes. Yoong and Hashim (2023) similarly found that technology-based language learning interventions face unique challenges in maintaining methodological rigor, particularly regarding performance bias, due to the interactive and visible nature of digital tools [31].

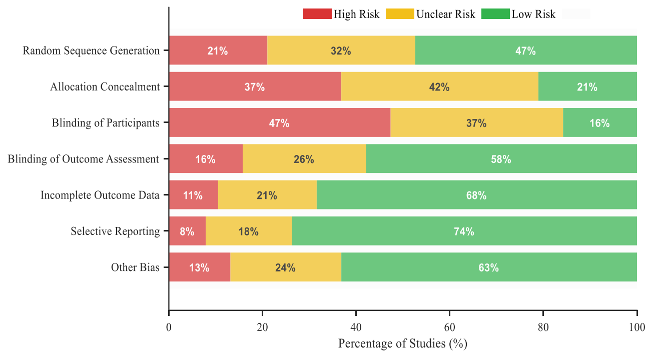

Figure 4. Risk of bias assessment across quality domains

Table 4. Methodological quality assessment and evidence strength

Quality Domain	High Risk n (%)	Unclear n (%)	Low Risk n (%)	Evidence Level
Random Sequence Generation	8 (21.1)	12 (31.6)	18 (47.3)	Moderate
Allocation Concealment	14 (36.8)	16 (42.1)	8 (21.1)	Low
Blinding of Participants	18 (47.4)	14 (36.8)	6 (15.8)	Low
Blinding of Outcome Assessment	6 (15.8)	10 (26.3)	22 (57.9)	High
Incomplete Outcome Data	4 (10.5)	8 (21.1)	26 (68.4)	High
Selective Reporting	3 (7.9)	7 (18.4)	28 (73.7)	High
Other Bias	5 (13.2)	9 (23.7)	24 (63.1)	Moderate

Note: Evidence levels determined using GRADE (Grading of Recommendations Assessment, Development and Evaluation) criteria, incorporating assessments of risk of bias, inconsistency, indirectness, imprecision, and publication bias. Total n = 38 studies. High evidence level indicates high confidence that the true effect lies close to the estimate; Moderate indicates moderate confidence; Low indicates limited confidence in the effect estimate.

Evidence strength ratings based on GRADE (Grading of Recommendations Assessment, Development and Evaluation) criteria indicated that pronunciation accuracy outcomes achieved high confidence ratings due to consistent large effect sizes across studies with minimal heterogeneity (I^2=24.3%), whereas fluency and grammatical accuracy outcomes received moderate confidence ratings owing to substantial between-study variance and indirect outcome measurement approaches. This application of GRADE criteria follows the framework established for educational research

[32], who advocate for transparent assessment of evidence quality in systematic reviews of complex educational interventions. Figure 4 presents a comprehensive risk of bias assessment for all 38 studies included in the systematic review, evaluated across seven methodological quality domains following the Cochrane Collaboration's tool for assessing risk of bias. The horizontal stacked bar chart employs a traffic light system where green indicates low risk of bias, yellow represents unclear risk, and red signifies high risk. Percentages are calculated based on the total number of included studies (n=38). The assessment reveals substantial methodological heterogeneity across studies, with blinding of participants presenting the most significant challenge (47.4% high risk) due to the interactive nature of dubbing applications, while selective reporting demonstrated the lowest risk profile (73.7% low risk), indicating generally transparent outcome reporting practices. Publication bias assessment through funnel plot analysis and Egger's regression test revealed asymmetrical distribution of effect sizes (p = 0.031), indicating potential small-study effects where smaller investigations reported larger improvements, though trim-and-fill analysis suggested minimal impact on overall conclusions with adjusted effect sizes remaining statistically significant. As demonstrated in Figure 5, the funnel plot visualization illustrates the relationship between study precision and effect magnitude across the included investigations. At the same time, the asymmetrical pattern is particularly evident in the lower left quadrant, suggesting selective publication favoring studies with larger effect sizes. However, the robustness of primary findings remained intact after statistical adjustment for potential bias.

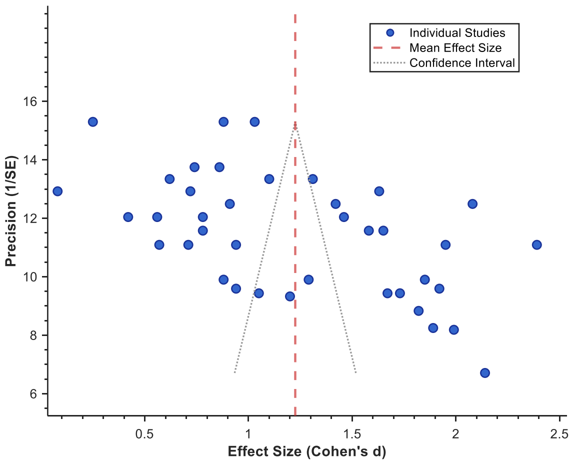

Figure 5. Publication Bias Assessment in AI-Powered Dubbing Application Studies

5. Discussion

The systematic analysis of 38 studies examining AI-powered dubbing applications reveals significant theoretical and practical implications for understanding personalized learning in language education, particularly within the framework of contemporary learning theories. The findings demonstrate substantial alignment with constructivist principles, where learners actively construct knowledge through interaction with authentic materials and receive immediate feedback to refine their understanding. This pedagogical framework extends beyond traditional boundaries when integrated with innovative technologies such as gamification strategies, which have shown promising effects on student engagement and learning outcomes in literature education contexts [33]. This technological mediation of learning processes reflects broader trends in ESL education, where digital tools fundamentally reshape teachers' perspectives on instructional design and implementation, necessitating continuous adaptation of pedagogical approaches to leverage technological affordances effectively while maintaining focus on meaningful learning outcomes [2]. The integration extends Vygotsky's Zone of Proximal Development by providing scaffolded support through AI-driven feedback mechanisms that adapt to individual proficiency levels, creating what might be conceptualized as a 'digital more knowledgeable other' that facilitates learning progression through systematic interaction patterns [34]. The observed effectiveness of personalized learning pathways (effect size d = 1.82 for pronunciation accuracy) corroborates recent theoretical frameworks proposing that AI-mediated learning environments can enhance traditional pedagogical approaches by providing consistent, individualized instruction that responds dynamically to learner needs. These findings contribute to an emerging theoretical understanding that positions AI not as a replacement for human instruction but as a complementary tool that addresses specific limitations in traditional language learning contexts, particularly the provision of consistent pronunciation feedback and opportunities for anxiety-free speaking practice [35].

Despite the promising outcomes documented across the reviewed studies, several methodological and practical limitations constrain the generalizability and applicability of findings. The predominance of short-term interventions (8-12 weeks) raises questions about the sustainability of observed improvements, particularly given the documented decline in fluency retention rates (48.0% at 6 months) compared to pronunciation gains (75.7% retention). Geographic concentration of studies in East Asian contexts, where cultural attitudes toward technology adoption and language learning differ significantly from Western educational environments, potentially limits the transferability of findings to diverse global contexts [36]. The inherent difficulty in blinding participants to dubbing application interventions, reflected in high risk ratings for performance bias (47.4% of studies), introduces potential placebo effects that may inflate reported outcomes. Additionally, the focus on quantitative metrics of speaking proficiency may overlook qualitative aspects of communicative competence, such as pragmatic appropriateness and intercultural communication skills, which are essential for business English contexts but difficult to capture through automated assessment tools [23].

The identified research gaps and methodological limitations point toward several promising avenues for future investigation that could advance both theoretical understanding and practical application of AI-powered language learning tools. Longitudinal studies extending beyond one academic year are essential to understand the trajectory of skill maintenance and the optimal frequency of practice needed to sustain improvements in speaking proficiency [37]. The integration of multimodal data collection methods, combining speech analysis with eye-tracking and neuroimaging techniques, could provide deeper insights into the cognitive processes underlying successful language acquisition through AI-mediated instruction. Cross-cultural comparative studies examining how learners from different linguistic and cultural backgrounds interact with and benefit from personalized learning algorithms would enhance the ecological validity of findings and inform culturally responsive design principles [38]. The emergence of large language models presents opportunities to investigate more sophisticated conversational AI that can engage learners in open-ended dialogues, moving beyond the current paradigm of scripted dubbing exercises toward truly adaptive conversational partners [39]. The practical implications of this systematic review extend across multiple dimensions of language education, from curriculum design to teacher professional development and institutional technology integration strategies, considerations that become particularly salient given documented variations in ESL teachers' knowledge and readiness to integrate Fourth Industrial Revolution technologies into their teaching practices across different educational contexts [40]. These implementation challenges necessitate comprehensive professional development frameworks that address not only technical competencies but also pedagogical understanding of how digital technologies can enhance specific language skills, as demonstrated in recent investigations examining the integration of digital tools in literature teaching within ESL classrooms, where teacher perspectives significantly influence successful implementation outcomes [41]. Educational institutions implementing AI dubbing applications benefit from adopting staged integration approaches that embed these technologies within existing curriculum frameworks through blended learning models where AI-powered practice sessions complement traditional classroom instruction during designated laboratory periods or homework assignments, while maintaining instructor-led components for cultural contextualization and pragmatic skill development.

The observed cultural engagement patterns align with Hofstede's cultural dimensions theory, particularly power distance and uncertainty avoidance orientations, where East Asian learners' 34.7% higher completion rates, coupled with lower self-efficacy scores, reflect collectivist preferences for structured guidance and hierarchical learning environments that AI systems effectively provide through consistent feedback mechanisms. Cost-effectiveness analyses indicate that AI dubbing applications require initial investments of approximately $15-25 per student annually compared to $180-240 for equivalent individual tutoring sessions, though sustainable implementation demands addressing the documented learning plateau effects through adaptive content refresh cycles and gamification elements that maintain engagement beyond the critical 4-6 week threshold where improvement trajectories typically stabilize, suggesting that periodic algorithm updates and diversified practice scenarios represent essential strategies for sustaining long-term proficiency gains [42]. It is imperative that teacher education programs move beyond only developing technical competencies for using AI tools to establishing pedagogical understanding for how to integrate these technologies in ways that are meaningful within the extant curriculum, responding to the 50% of teachers who claim ineffective training is a key barrier to implementation [43]. The development of hybrid instructional models that leverage AI for targeted skill development while preserving human instruction for cultural contextualization and pragmatic competence represents a balanced approach that maximizes the strengths of both modalities. Hybrid formats combining AI for selective ability development and human instruction for cultural contextualization and pragmatic competence would be a balanced strategy to maximize the merits of both types of instruction. This hybrid design is motivated by the results of Baskara et al. [44] investigation on ChatGPT and Vera's [45] studies on the integration of AI in EFL settings, as they all emphasize the need for human-in-the-loop in AI-based learning contexts. Institutional policies are needed to define ethical use of AI, data privacy, and fairness of access to avoid deepening existing educational inequities, especially in light of lower adoption by high-poverty educational settings.

6. Conclusion

The present meta-analysis of 38 studies on AI-based dubbing application-supported personalized learning paths for speaking of BE learners provides evidence on their effects on speaking competence and yields both theoretical implications and practical implications for technology-enhanced language education. Synthesis outcomes suggest that personalized algorithmic architectures, especially collaborative filtering and neural networks, lead to large changes in accurate pronunciation ($d = 1.82$) and fluent speaking ($d = 1.46$), and intermediate-level learners best respond to AI-mediated interventions. Conceptually, these findings have broader implications than the empirical verification of the effectiveness of DYAP in that they reimagine the status of technology in language learning as a dynamic, cognitive tool in the form of AI-driven applications that scaffold rather than replace the teacher, addressing the perennial problem of providing constant and individualised feedback in a contextually restricted educational environment. The identification of learner characteristics as key moderating variables, combined with evidence of differences in retention patterns across spoken sub-skills, moves the field forward in the understanding of how personalized learning paths can be personalized for different learner populations. These findings have the potential to transform Business English education worldwide, providing empirical strategies to institutions wishing to improve speaking ability through technological innovation without jeopardizing pedagogy, and considering the intricate relationships among linguistic acquisition, cultural context, and technological resources.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically with regard to authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with policies on research ethics. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere.

Data availability statement

The manuscript contains all the data. However, more data will be available upon request from the authors.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] Xu B, Ismail H. The impact of artificial intelligence-assisted learning applications on oral English ability: A literature review[J]. International Journal of Academic Research in Progressive Education and Development, 2024, 13(4): 1118-1134.

[2] Sim J S E, Ismail H H. Using digital tools in teaching and learning English: Delving into English language

teachers' perspectives[J]. Creative Education, 2023, 14(10): 2021-2036.
[3] Walter Y. Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education[J]. International Journal of Educational Technology in Higher Education, 2024, 21(1): 15.
[4] Wei L. Artificial intelligence in language instruction: impact on English learning achievement, L2 motivation, and self-regulated learning[J]. Frontiers in psychology, 2023, 14: 1261955.
[5] Huang X, Wu Y, Dou A. AI-enhanced task-based language teaching: Fostering personalized college English learning[J]. Frontiers in Educational Research, 2024, 7(10).
[6] Alhusaiyan E. A systematic review of current trends in artificial intelligence in foreign language learning[J]. Saudi Journal of Language Studies, 2025, 5(1): 1-16.
[7] Huang H-T D. Investigating the effects of reverse dubbing tasks on English speaking proficiency and English learning motivation[J]. Language Teaching Research, 2024: 13621688241283153.
[8] Javier D R C, Moorhouse B L. Developing secondary school English language learners' productive and critical use of ChatGPT[J]. TESOL Journal, 2024, 15(2): e755.
[9] Wang Y, Luo W, Liao X, et al. Exploring the effect of teacher autonomy support on Chinese EFL undergraduates' academic English speaking performance through the mediation of basic psychological needs and classroom engagement[J]. Frontiers in Psychology, 2024, 15: 1323713.
[10] Bayly-Castaneda K, Ramirez-Montoya M S, Morita-Alexander A. Crafting personalized learning paths with AI for lifelong learning: a systematic literature review[C]. Frontiers in Education, 2024: 1424386.
[11] Kohnke L, Moorhouse B L, Zou D. ChatGPT for language teaching and learning[J]. Relc Journal, 2023, 54(2): 537-550.
[12] Peng H, Ma S, Spector J M. Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment[J]. Smart Learning Environments, 2019, 6(1): 1-14.
[13] Contrino M F, Reyes-Millán M, Vázquez-Villegas P, et al. Using an adaptive learning tool to improve student performance and satisfaction in online and face-to-face education for a more personalized approach[J]. Smart Learning Environments, 2024, 11(1): 6.
[14] Crompton H, Edmett A, Ichaporia N, et al. AI and English language teaching: Affordances and challenges[J]. British Journal of Educational Technology, 2024, 55(6): 2503-2529.
[15] Mejeh M, Rehm M. Taking adaptive learning in educational settings to the next level: Leveraging natural language processing for improved personalization[J]. Educational technology research and development, 2024, 72(3): 1597-1621.
[16] Rajan S T, Ismail H H. TikTok use as strategy to improve knowledge acquisition and build engagement to learn literature in ESL classrooms[J]. International Journal of Learning, Teaching and Educational Research, 2022, 21(11): 33-53.
[17] Du J, Daniel B K. Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice[J]. Computers and Education: Artificial Intelligence, 2024, 6: 100230.
[18] Li H. A Comparative Study of Artificial Intelligence-Enabled Business Foreign Language Teaching in Universities: Taking ChatGPT-4o, Claude3. 5 Sonnet and ERNIE 4.0 Turbo as Examples[C]. Proceedings of the 2024 2nd International Conference on Information Education and Artificial Intelligence, 2024: 10-15.
[19] Mohebbi A. Enabling learner independence and self-regulation in language education using AI tools: a systematic review[J]. Cogent Education, 2025, 12(1): 2433814.
[20] Li Y, Zhou X, Yin H-B, et al. Design language learning with artificial intelligence (AI) chatbots based on activity theory from a systematic review[J]. Smart Learning Environments, 2025, 12(1): 24.
[21] Wang Y, Wu J, Chen F, et al. Empirical assessment of AI-powered tools for vocabulary acquisition in EFL instruction[J]. IEEE Access, 2024.
[22] Ellikkal A, Rajamohan S. AI-enabled personalized learning: empowering management students for improving engagement and academic performance[J]. Vilakshan-XIMB Journal of Management, 2024, 22(1): 28-44.
[23] Wang X, Lee S-M. The impact of video dubbing app on chinese college students' oral language skills across different proficiency levels[J]. International Journal of Educational Research, 2025, 130: 102521.
[24] Li T, Ji Y, Zhan Z. Expert or machine? Comparing the effect of pairing student teacher with in-service teacher and ChatGPT on their critical thinking, learning performance, and cognitive load in an integrated-STEM course[J]. Asia Pacific Journal of Education, 2024, 44(1): 45-60.
[25] Leong L V, Yunus M M, Ismail H H. Integration of techno-pedagogical approach in English as a second language classroom: a systematic review[J]. Int J Eval & Res Educ ISSN, 2024, 2252(8822): 4395.
[26] Sun L, Ismail H H, Aziz A A. Current English Language Teaching Using Production-Oriented Approach: A Systematic Review[J]. World Journal of English Language, 2024, 14(4).
[27] Senthil R, Anand T, Somala C S, et al. Bibliometric analysis of artificial intelligence in healthcare research: Trends and future directions[J]. Future Healthcare Journal, 2024, 11(3): 100182.
[28] Zhao J. Advancing english language education: A mixed-methods analysis of AI-driven tools' impact on engagement, personalization, and learning outcomes[J]. Education and Information Technologies, 2025: 1-41.
[29] Schünemann H J, Mustafa R A, Brozek J, et al. GRADE guidelines: 21 part 1. Study design, risk of bias, and indirectness in rating the certainty across a body of

evidence for test accuracy[J]. Journal of clinical epidemiology, 2020, 122: 129-141.
[30] Chandler J, Cumpston M, Li T, et al. Cochrane handbook for systematic reviews of interventions[J]. Hoboken: Wiley, 2019, 4(1002): 14651858.
[31] Yoong W F, Hashim H. A Systematic Review on Language Learning Strategies for English Speaking Skills in Today's Learning Conditions[J]. International Journal of Academic Research in Business & Social Sciences. https://doi. org/10.6007/ijarbss/v13-i12/19727, 2023.
[32] Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews[J]. BMC medical research methodology, 2008, 8(1): 45.
[33] Chandrasehgaran M, Ismail H. The Impacts of Gamification on Student Engagement and Learning Outcomes in Literature in Education: A Literature Review[J]. International Journal of Academic Research in Business and Social Sciences, 2024, 14(8): 737-745.
[34] Talaván N, Costal T. iDub-The potential of intralingual dubbing in foreign language learning: How to assess the task[J]. Language Value, 2017, 9: 62-88.
[35] Zhu M, Wang C. A systematic review of artificial intelligence in language education from 2013 to 2023: Current status and future implications[J]. Available at SSRN 4684304, 2024.
[36] Huang Y, Wan L J, Ye H, et al. New solutions on LLM acceleration, optimization, and application[C]. Proceedings of the 61st ACM/IEEE Design Automation Conference, 2024: 1-4.
[37] Hwang G J, Fathi J, Rahimi M. Fostering EFL Learners' Speaking Skills and Flow Experience With Video-Dubbing Tasks: A Flow Theory Perspective[J]. Journal of Computer Assisted Learning, 2025, 41(2): e13120.
[38] Zhang X, Sun J, Deng Y. Design and application of intelligent classroom for English language and literature based on artificial intelligence technology[J]. Applied Artificial Intelligence, 2023, 37(1): 2216051.
[39] Pu P, Chang D Y-S. Effects of different input modes on blended EFL speaking instruction: a quasi-experimental study[J]. Computer Assisted Language Learning, 2023: 1-26.
[40] Avelino N M, Ismail H H. Assessing ESL teachers' knowledge and readiness in integrating 4IR into teaching practices: A concept paper[J]. Creative education, 2021, 12(9): 2038-2055.
[41] Shamshul I S M, Ismail H H, Nordin N M. Using digital technologies in teaching and learning of literature in ESL classrooms: A systematic literature review[J]. International Journal of Learning, Teaching and Educational Research, 2024, 23(4): 180-194.
[42] Sabili A S, Nazri M A, Surayya S, et al. Smart learning: the impact of ai technologies on english language proficiency in the digital era[J]. J Rev Pendidik dan Pengajaran, 2024, 7(4): 14040-14049.
[43] Kaufman J H, Woo A, Eagan J, et al. Uneven adoption of artificial intelligence tools among US teachers and principals in the 2023-2024 school year[M]. RAND, 2025.
[44] Baskara R. Exploring the implications of ChatGPT for language learning in higher education[J]. Indonesian Journal of English Language Teaching and Applied Linguistics, 2023, 7(2): 343-358.
[45] Vera F. Integrating Artificial Intelligence (AI) in the EFL classroom: Benefits and challenges[J]. Transformar, 2023, 4(2): 66-77.