REVIEW OF ARTIFICIAL INTELLIGENCE IN EDUCATION

DOI: https://doi.org/10.37497/rev.artif.intell.educ.v7ii.1120

ARTICLE

EDUCATIONAL TECHNOLOGIES FOR MULTILINGUAL LEARNERS: A SYSTEMATIC REVIEW OF AI-BASED HUMAN-CENTERED DESIGN

Tecnologias Educacionais para Aprendizes Multilíngues: Uma Revisão Sistemática de Design Centrado no Humano Baseado em Inteligência Artificial

Taraneh Yarahmadi
Department of Design, Politecnico di Milano, Via Durando 38/A, 20158, Milan, Italy.
Email: taraneh.yarahmadi@mail.polimi.it

Zane Whitcomb
Department of Sociology, California State University, Sacramento, 6000 J Street, Sacramento, 95819, CA, USA.
Email: zwhitcomb@csus.edu

ABSTRACT | Background: While AI is promoted as a transformative force in education through adaptive platforms and real-time feedback, its implementation for multilingual learners often risks reinforcing educational inequalities. Current scholarship cautions that automated systems still struggle with cultural nuances and idiomatic expressions, highlighting the need for design approaches that foreground equity and human agency. **Objective:** This systematic review examines 10 core studies through the lens of the ISO 9241-210 Human-Centered Design framework. The objective is to analyze how AI-based educational technologies are designed and evaluated to support multilingual learners, specifically focusing on the "Context of Use," "User Requirements," "Design Solutions," and "Evaluation" phases. **Methods:** Through a systematic search of four databases and subsequent snowballing, 10 core papers were selected to analyze how AI tools address linguistic and cultural diversity. **Results:** Across the reviewed studies, reported improvements ranged from quantitative gains, including a 25–31% increase in literacy and vocabulary retention and a rise in academic success rates up to 77.8%. Furthermore, AI-driven systems, when aligned with HCD principles, were associated with saving educators up to 41% of their time. However, evaluations revealed critical 'socio-technical paradoxes': students faced a "trade-off" where they reverted to English-centric prompting due to algorithmic bias in low-resource languages, and risks of "metacognitive laziness" emerged from over-reliance on automated tools. **Conclusion:** The successful integration of AI in multilingual contexts depends on a shift from content generation to pedagogical scaffolding. Designers must prioritize "Human-in-the-loop" models that balance computational efficiency with human oversight to provide the emotional engagement and cultural authenticity that AI currently lacks.

Keywords | Human-Centered Design, ISO 9241-210, Multilingual Education, Artificial Intelligence, Natural Language Processing.

Qualis B1 CAPES 2021-2024

Indexed in ANPAD SPELL ANPAD

Received: 29 June 2026
Revised: 29 June 2026
Accepted: 05 August 2026
e-ISSN: 2965-4688

Corresponding Author:
Taraneh Yarahmadi – E-mail: taraneh.yarahmadi@mail.polimi.it

How to cite this article: Yarahmadi, T., & Whitcomb , Z. (2026). Educational Technologies for Multilingual Learners: A Systematic Review of AI-Based Human-Centered Design. *Review of Artificial Intelligence in Education,* 7(i), e01120. https://doi.org/10.37497/rev.artif.intell.educ.v7ii.1120

RESUMO | Contexto: Embora a inteligência artificial seja promovida como uma força transformadora na educação por meio de plataformas adaptativas e feedback em tempo real, sua implementação para aprendizes multilíngues frequentemente corre o risco de reforçar desigualdades educacionais. A literatura atual alerta que sistemas automatizados ainda apresentam dificuldades com nuances culturais e expressões idiomáticas, evidenciando a necessidade de abordagens de design que priorizem a equidade e a agência humana. **Objetivo:** Esta revisão sistemática examina 10 estudos centrais por meio da lente do framework de Design Centrado no Humano ISO 9241-210. O objetivo é analisar como tecnologias educacionais baseadas em IA são projetadas e avaliadas para apoiar aprendizes multilíngues, com foco específico nas fases de "Contexto de Uso", "Requisitos do Usuário", "Soluções de Design" e "Avaliação". **Métodos:** Por meio de busca sistemática em quatro bases de dados e subsequente técnica de snowballing, 10 artigos centrais foram selecionados para analisar como ferramentas de IA abordam a diversidade linguística e cultural. **Resultados:** Nos estudos revisados, as melhorias relatadas variaram desde ganhos quantitativos, incluindo aumento de 25–31% na retenção de literacia e vocabulário e elevação das taxas de sucesso acadêmico para até 77,8%. Além disso, sistemas orientados por IA, quando alinhados aos princípios de DUH, foram associados a uma economia de até 41% do tempo dos educadores. No entanto, as avaliações revelaram "paradoxos sociotécnicos" críticos: os estudantes enfrentaram um dilema no qual revertiam para prompts em inglês devido ao viés algorítmico em línguas de baixos recursos, e riscos de "preguiça metacognitiva" emergiram da dependência excessiva de ferramentas automatizadas. **Conclusão:** A integração bem-sucedida da IA em contextos multilíngues depende de uma mudança da geração de conteúdo para o scaffolding pedagógico. Designers devem priorizar modelos "Human-in-the-Loop" que equilibrem a eficiência computacional com a supervisão humana, de modo a proporcionar o engajamento emocional e a autenticidade cultural que a IA atualmente não é capaz de oferecer.

Palavras-chave | Design Centrado no Humano; Inteligência Artificial na Educação; Aprendizes Multilíngues; Viés Algorítmico; ISO 9241-210

INTRODUCTION

In recent years, artificial intelligence (AI) has been promoted as a transformative force in education through adaptive platforms and data-driven personalization.(Aravind et al., 2025; Shoeibi et al., 2023) Leveraging natural language processing (NLP) and generative architectures, these systems provide real-time feedback and adjust content difficulty to enhance digital learning (Aravind et al., 2025),(Shoeibi et al., 2023). However, scholarship underscores that benefits are not automatic; without attention to equity and human agency, AI risks reinforcing inequalities (Saddhono et al., 2024). Language learning has emerged as a central application area for these tools (Aravind et al., 2025).

Multilingual education is essential for cognitive development and intercultural competence (Saddhono et al., 2024), (Aravind et al., 2025). While approaches like immersion and bilingual instruction foster cultural empathy and problem-solving skills, access remains uneven for marginalized learners (Saddhono et al., 2024) . These disparities are salient as institutions confront growing linguistic diversity (Saini, 2025). AI is often positioned to address these challenges, with NLP-based tools reducing the cognitive load of navigating multiple languages and improving comprehension (Saini, 2025). Personalized AI systems have shown gains in vocabulary and syntax through gamified chatbots and targeted recommendations (Aravind et al., 2025). However, truly effective AI design must account for the Multilingual Age of Exposure, as the timing of language acquisition significantly influences how users interact with and internalize automated feedback. (Khan et al., 2025)

Despite this potential, research cautions against techno-solutionist narratives (Saddhono et al., 2024). AI models often suffer from a "*Linguistic Power Gap*"; because they are trained on English-centric datasets, they frequently default to a monolingual bias that erases local dialects and distorts cultural identity (Krstinić & Bačić, 2025). Automated systems still struggle with cultural nuance and idiomatic expressions (Saddhono et al., 2024), (Saini, 2025). Furthermore, algorithmic bias and data privacy concerns are acute for multilingual learners whose cultural practices are under-represented in training data (Krstinić & Bačić, 2025). Realizing AI's potential is thus a fundamental challenge of ***interaction design*** and socio-technical inclusivity (Schmager et al., 2025).

To move beyond standardized narratives, educational tools must be evaluated through the ***ISO 9241-210*** standard for Human-Centered Design (HCD). This framework shifts focus from machine output to the iterative cycle of understanding users, specifying requirements, and evaluating outcomes. Currently, there is limited synthesized evidence on how AI technologies for multilingual learners are evaluated from this human-centered perspective(Saini, 2025), (Aravind et al., 2025). This systematic review addresses this gap by analyzing AI-based technologies through the HCD lens.

REVIEW OBJECTIVE AND QUESTION

- **RQ1:** How do AI-based educational technologies for multilingual learners operationalize the *Context of Use* and *User Requirements*?
- **RQ2:** What *Design Solutions* and interaction models are implemented to support pedagogical goals?
- **RQ3:** How are these systems evaluated regarding both learning outcomes and socio-technical impacts like algorithmic bias?

METHODS

The selection of studies was governed by the following criteria to ensure a focused analysis of **Human-Centered Design (HCD)** in AI-supported multilingual education:

- **Participants:** Learners across all developmental stages, including early childhood (ages 5–12), secondary, higher education, and adult immigrant populations, as well as educators and pre-service teachers (PSTs).
- **Language Context:** Individuals engaged in second/foreign language acquisition (e.g., EFL, ELL, NNES) or bilingual/multilingual instructional programs.
- **Technology Scope:** Implementation of Artificial Intelligence, including Generative AI (ChatGPT, Gemini), NLP-based chatbots, AI-generated virtual avatars, and adaptive machine learning algorithms.
- **Application Areas:** Tools facilitating collaborative writing, peer-review support, or automated discourse analysis.
- **Educational Settings:** Both formal academic institutions (K-12 and higher education) and informal or non-traditional environments (e.g., newcomer integration programs).

- **Exclusion Criteria:** Research focusing exclusively on native monolingual learners or studies lacking a clear focus on language acquisition, bilingualism, or literacy development.

SEARCH STRATEGY AND STUDY SELECTION

The systematic search was conducted across four primary academic databases (Scopus, Web of Science, Sage, and ACM Digital Library), supplemented by Taylor & Francis (T&F) and Google Scholar for backward and forward "snowballing".

- **Scopus:** TITLE-ABS-KEY("human-centered design" OR "user-centered design" OR "inclusive design" OR "participatory design" OR "empathic design") AND TITLE-ABS-KEY("artificial intelligence" OR "AI" OR "generative AI" OR "natural language understanding" OR "conversational AI") AND TITLE-ABS-KEY("educational technology" OR "personalized learning" OR "adaptive learning systems" OR "interactive learning tools") AND TITLE-ABS-KEY("multilingual" OR "bilingual" OR "language diversity" OR "English language learners" OR "language-inclusive design")
- **Web of Science (WoS) + Sage:** TS=("human-centered design" OR "user-centered design" OR "inclusive design" OR "participatory design" OR "empathic design") AND TS=("artificial intelligence" OR "AI" OR "generative AI" OR "natural language understanding" OR "conversational AI") AND TS=("educational technology" OR "personalized learning" OR "adaptive learning systems" OR "interactive learning tools") AND TS=("multilingual" OR "bilingual" OR "language diversity" OR "English language learners" OR "language-inclusive design")
- **ACM Digital Library:** ("human-centered design" OR "user-centered design" OR "inclusive design" OR "participatory design" OR "empathic design") AND ("artificial intelligence" OR "AI" OR "generative AI" OR "natural language understanding" OR "conversational AI") AND ("educational technology" OR "personalized learning" OR "adaptive learning systems" OR "interactive learning tools") AND ("multilingual" OR "bilingual" OR "language diversity" OR "English language learners" OR "language-inclusive design")

  Following the initial database search which identified 34 records, Google Scholar and Taylor & Francis (T&F) were utilized for snowballing. This involved:

- **Backward Snowballing:** Reviewing the reference lists of the identified 34 records to find earlier foundational HCD studies.
- **Forward Snowballing:** Identifying newer studies that cited your core papers (like Saddhono et al., 2024) to ensure the review captured the most current 2025–2026 developments.

Data Extraction

The systematic review followed a structured collaborative protocol to enhance objectivity, followed by a primary-led synthesis phase. Initially, two reviewers performed a split screening of the ***34 identified unique*** records to ensure cross-verification of eligibility based on the pre-defined inclusion and exclusion criteria.

Following this initial screening, the primary author conducted the final, in-depth analysis of the 10 selected studies, specifically extracting and synthesizing the Human-Centered Design (HCD) components against the ISO 9241-210 framework. This phase involved the primary-led consolidation of quantitative data (e.g., literacy improvements) and qualitative insights into a unified thematic structure. Furthermore, the primary author was responsible for the development of all analytical charts, including the PRISMA flow diagram and the hierarchical HCD process models, to visually communicate the core findings of the review. To ensure inter-rater reliability, the second author performed a validation check on 20% of the extracted data, with any minor discrepancies resolved through iterative discussion.

Evaluation Framework

To systematically assess the human-centered maturity and pedagogical efficacy of the identified AI tools, this review employed a multi-dimensional evaluation framework grounded in the ***ISO 9241-210:2019*** standard (Zhang, 2024). Following the non-empirical evaluation methodology proposed by ***Paniagua et al.*** , each study was mapped against the four iterative phases of the Human-Centered Design (HCD) lifecycle: (1) understanding the context of use, (2) specifying user requirements, (3) producing design solutions, and (4) evaluating the design against requirements (Zhang, 2024).

Furthermore, the analysis incorporated the comprehensive evaluation dimensions identified by Sargent and Calderón(Youness Moudettir et al., 2025) specifically focusing on cognitive ergonomics and the pedagogical quality of the interaction models. This combined approach allowed for a critical synthesis of how AI-driven interventions for multilingual learners balance computational efficiency with human agency and socio-technical inclusivity.

Study selection

The selection process: The initial search yielded **34 records**, which was reduced to **30** after removing duplicates. Following a title and abstract screening, **17 full-text articles** were assessed for eligibility. Ultimately, **10 studies** were retained for this review.

The final selection includes a diverse range of research designs:

- **Experimental & Mixed-Methods:** Used to measure literacy gains, vocabulary retention, and syntax accuracy.
- **Qualitative & Case Studies:** Explored student narratives, classroom implementation, and discourse analysis.
- **Comparative & Empirical Studies:** Analyzed cross-cultural AI adoption and the efficacy of multilingual prompting in technical subjects.
- **Design-Based & Deployment Studies:** Investigated “in-the-wild” tool usage and pedagogical interventions to critique algorithmic bias.

Prisma

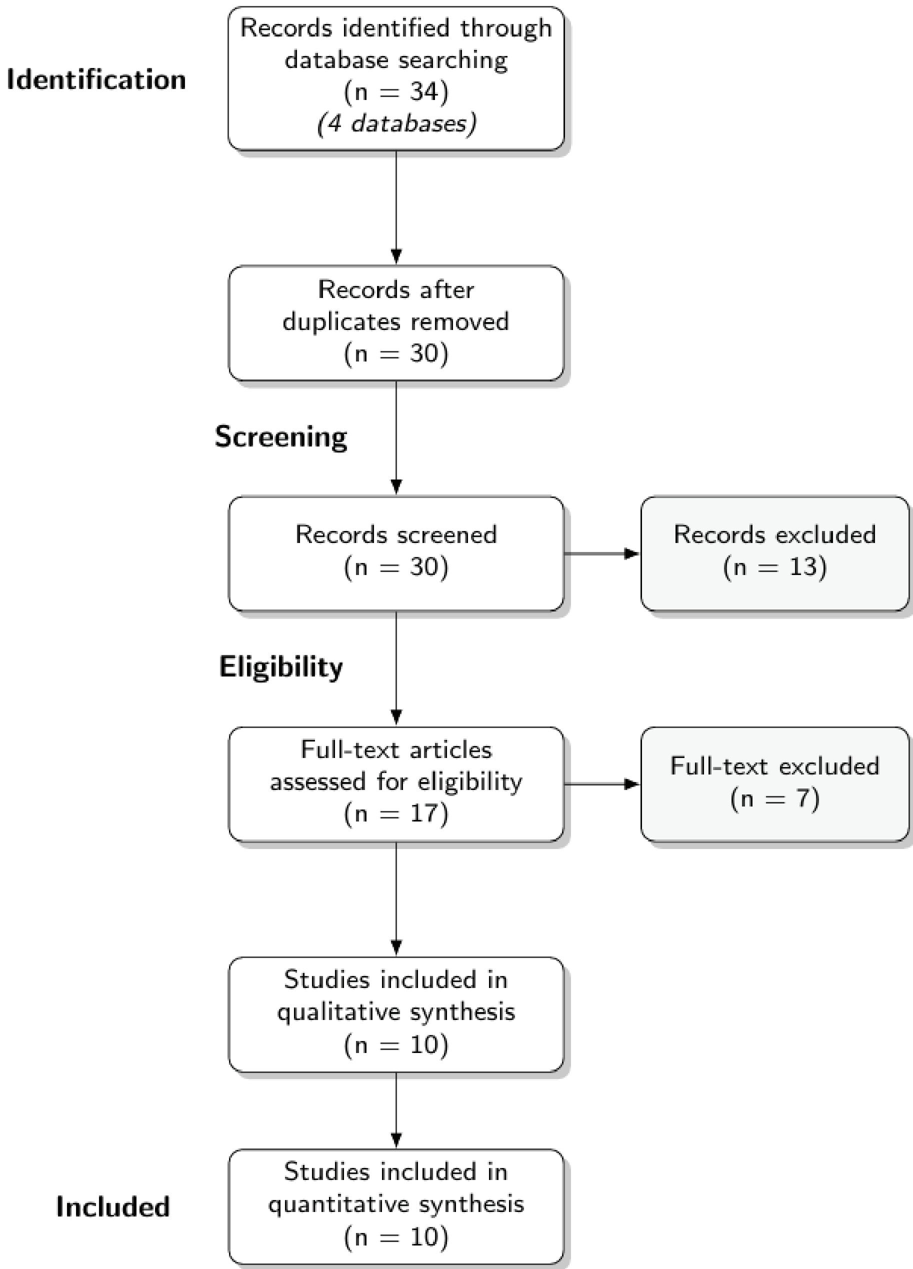

Figure 1. PRISMA Flow Diagram of the Study Selection Process

Study Characteristics

1. **Aravind et al. (2025)**(Aravind et al., 2025) This experimental study investigated an Adaptive AI-Based Personalized Learning System designed for young English learners (ages 5–12). The intervention integrated Natural Language Processing (NLP) chatbots, machine learning algorithms, and gamification to create dynamic, personalized learning paths. Using a pre-

test/post-test design with an experimental and a control group, the researchers measured the system's impact on accelerating vocabulary retention and syntax accuracy compared to traditional teaching methods.

2. **Cole (2025)**(Cole, 2025) Set in a bilingual secondary school in Paris, France, this case study focused on 16 to 17-year-old students preparing for high-stakes exams. The research explored the integration of an AI tool called "Classroom Companion" into the curriculum. The study characteristics highlight the use of AI to provide immediate, rubric-based feedback on timed writing assignments, aiming to enhance students' written expression skills, reinforce basic grammar, and reduce teacher grading workloads.
3. **Khan et al. (2025)**(Khan et al., 2025) This qualitative study explored the lived experiences of 43 Indian English as a Foreign Language (EFL) learners in higher education who actively use ChatGPT for academic writing. Grounded in Cognitive Load Theory and the Technology Acceptance Model, the researchers utilized semi-structured interviews to examine how ChatGPT impacts writing fluency, learner confidence, time management, and cognitive engagement, while also highlighting concerns regarding academic integrity and over-reliance.
4. **Konca et al. (2025)**(Konca et al., 2025) A comparative quantitative study that analyzed the cross-cultural adoption of AI among 846 pre-service teachers in Turkey and the United Arab Emirates (UAE). Utilizing the GETAMEL (General Extended Technology Acceptance Model for E-Learning) framework, the researchers employed Structural Equation Modeling (SEM) to determine how cultural contexts and factors like subjective norms, self-efficacy, and anxiety influence behavioral intentions to use AI in educational practices.
5. **Liaqat & Munteanu (2020)**(Liaqat & Munteanu, 2020) This deployment study focused on 16 mature immigrant English Language Learners (ELLs) engaged in informal learning contexts (the Language Instruction for Newcomers to Canada program). Through formative research and "in-the-wild" deployment, the researchers designed a web-based, community-centered peer-review application. The study evaluated how the use of localized, tagged rubrics could elicit relevant formative feedback and provide sustainable writing support without an instructor present.
6. **Miranda & Vegliante (2025)**(Miranda & Vegliante, 2025) An experimental and comparative study involving 147 participants from the European mining sector across Türkiye, Poland, Portugal, and Italy. The research evaluated the effectiveness of integrating AI-generated virtual speakers (created via the ELAI platform) into multilingual e-learning courses (e.g., mine rescue, mental wellbeing). The study assessed learner preferences for AI avatars speaking their native languages versus human speakers delivering content with subtitles.
7. **Prather et al. (2025)**(Prather et al., 2024) A multi-site case study involving approximately 80 non-native English speaking computer science students in Portugal, New Zealand, and Saudi Arabia. The study investigated the efficacy of multilingual prompting by having students use Generative AI (via a tool called "Promptly") to solve programming tasks ("Prompt Problems") in their native languages (Portuguese, Chinese, Arabic. Both quantitative success rates and qualitative user experiences were analyzed.

8. **Rivero & Yin (2025)**(Rivero & Yin, 2025) A qualitative, design-based intervention examining the perceptions of 13 bilingual pre-service teachers (Mandarin-English and Spanish-English) in a US graduate-level teacher preparation program. Through reflective journaling and the design of AI-generated multimodal texts, the study explored how participants navigated the tension between AI's creative affordances for bilingual education and its constraints, specifically how AI algorithms often reproduce standardized English and monolingual ideologies.
9. **Saddhono et al. (2024)**(Saddhono et al., 2024) A mixed-methods study that evaluated the impact of an integrated AI ecosystem on 500 students from diverse language backgrounds over one academic year. The study tracked the implementation of AI-driven language learning apps (like Duolingo and Babbel), automated translation tools (like Google Translate), and cultural education platforms. Outcomes measured included multilingual literacy scores, translation accuracy, cultural understanding, and overall student engagement.
10. **Zhang (2024)**(Zhang, 2024) A qualitative case study featuring an in-depth interview with a bilingual educator managing a 5th-grade Chinese-English classroom in Arizona, USA. The research explored the practical implementation of AI tools (ChatGPT, Gemini, Magic School, Canva) for lesson planning, creating gamified activities, and specifically for conducting discourse analysis. The study highlighted how AI assists educators in understanding the depth of young children's bilingual responses and providing targeted feedback.

Table 1. Summary of Included Studies

Study	Project	Setting	Design Goal	Participants	Methods
Aravind et al. (2025)	Adaptive AI-Based Personalized Learning System	Educational institutions (Schools/ Language Training Centers)	Develop an AI-driven system using NLP chatbots and gamification to accelerate vocabulary recall and syntax accuracy.	Young English Learners (aged 5–12). *N* not specified, divided into experimental and control groups.	**Experimental Study:** (1) **Pre-Test:** Participants underwent an initial assessment (X1) measuring baseline vocabulary and syntax proficiency via MCQs and oral tasks. (2) **Intervention Phase:** The *Experimental Group* used the AI-powered system (NLP chatbots, adaptive models, gamification), while the *Control Group* used traditional methods (lectures, worksheets). Data on engagement and learning pace were tracked continuously. (3) **Post-Test:** A final assessment (X2) measured retention and comprehension. (4) **Qualitative Feedback:** Surveys and interviews were conducted to assess usability and engagement using thematic and sentiment analysis.
Cole (2025)	Using AI (Classroom Companion) to Enhance Writing Skills	Secondary School (Collège Sévigné, Paris, France); Bilingual Section	Integrate AI tools to improve written expression, reinforce grammar/vocabulary, and manage teacher workload.	16 to 17-year-old language learners (11th grade/Junior year). Two classes.	**Case Study:** (1) **Implementation:** Teachers integrated "Classroom Companion" for bi-monthly assignments. (2) **Activity I:** Students submitted responses to a prompt (e.g., on *Intimate Apparel*) with a 20-minute limit. The AI provided immediate feedback based on a rubric; students had 3 chances to revise and resubmit. (3) **Activity II:** A follow-up timed activity (10 short responses in 45 min) was conducted 4 weeks later. (4) **Evaluation:** Success rates, response times, and feedback quality were analyzed (comparing Jan to Feb 2025 data). Qualitative observations of student/teacher reactions were recorded.

Study	Project	Setting	Design Goal	Participants	Methods
Khan et al. (2025)	Exploring Indian EFL learners' academic writing narratives with ChatGPT	Higher Education (Various universities in India)	Investigate the impact of ChatGPT on writing fluency, learner confidence, and cognitive engagement using CLT and TAM frameworks.	43 Indian EFL learners (21 female, 22 male) who have used ChatGPT for >1 year.	**Qualitative Study:** (1) **Recruitment:** Purposive sampling was conducted via Google Forms to identify experienced users. (2) **Data Collection:** Semi-structured interviews (25–30 mins) were conducted virtually via Google Meet. Questions focused on usability, engagement, and feedback utility. (3) **Analysis:** Thematic analysis was performed using Clarke and Braun's six-step framework to identify key themes (e.g., Content Quality, Efficiency, Ethical Ambiguity).
Konca et al. (2025)	Cross-cultural perspectives on AI adoption in teacher education	Faculties of Education in Turkey and the United Arab Emirates (UAE)	Compare factors influencing pre-service teachers' behavioral intention to use AI using the GETAMEL framework (TAM extension).	846 Pre-service teachers (406 from Turkey, 440 from UAE).	**Comparative Quantitative Study:** (1) **Instrument Design:** A survey was developed based on GETAMEL constructs (Subjective Norm, Experience, Anxiety, Self-Efficacy, etc.) and translated into Turkish and Arabic. (2) **Pilot Study:** Conducted with 134 (Turkey) and 142 (UAE) participants to validate the instrument. (3) **Data Collection:** The final survey was disseminated via online university portals. (4) **Analysis:** Confirmatory Factor Analysis (CFA) and Structural Equation Modeling (SEM) were used to test hypotheses and compare model fit between the two countries.
Liaqat & Munteanu (2020)	Leveraging Peer Support for Mature Immigrants Learning to Write	Informal Contexts (Language Instruction for Newcomers to Canada - LINC)	Design a community-centered peer-review tool to support writing development for mature immigrant learners.	16 mature immigrant English Language Learners (ELLs).	**Deployment Study:** (1) **Formative Phase:** Prior analysis of writing samples and participatory design sessions established design requirements. (2) **App Deployment:** Participants used a web app for 9 days to complete 3 writing assignments. (3) **Peer Review:** Users engaged in a 4-step process: Submit, Tag (classify parts of their text), Review Peer (using localized vs. general rubrics), and Revise. (4) **Evaluation:** Writing samples and revisions were analyzed for accuracy. (5) **Focus Groups:** Three focus groups (semi-structured interviews) were conducted to discuss the writing process, feedback trust, and community sustainability.
Miranda & Vegliante (2025)	DigiRescueMe: AI-Generated Virtual Speakers in E-Learning	Educational/ Vocational contexts (Mining sector focus) in Türkiye, Poland, Portugal, Italy.	Enhance accessibility and engagement in multilingual e-learning using AI-generated virtual speakers vs. human speakers.	147 participants (Students, Researchers, Technicians, etc.).	**Experimental/Comparative Study:** (1) **Course Development:** Three Moodle courses were created; two used AI virtual speakers (Experimental Group - EG), one used a human speaker (Control Group - CG). (2) **Intervention:** Participants selected a course and completed it asynchronously between Nov–Dec 2024. (3) **Assessments:** Participants completed mandatory initial and final multiple-choice assessments. (4) **Questionnaire:** A post-course survey (18 questions, based on UEQ/TUXEL) collected data on satisfaction, preference for native language content, and perception of AI vs. human speakers.
Prather et al. (2025)	Breaking the Programming Language Barrier (Multilingual Prompting)	Higher Education (Universities in Portugal, New Zealand, Saudi Arabia)	Explore the efficacy of using non-English native languages to solve "Prompt Problems" (GenAI coding tasks).	~80 Students (27 Portuguese, 19 Chinese speakers, 34 Arabic speakers).	**Multi-site Case Study:** (1) **Task Performance:** Students attempted to solve "Prompt Problems" (generating code via GenAI) using their native languages. (2) **Data Collection:** 1,679 prompts were collected and categorized by linguistic strategy (Native, English, Mixed). (3) **Quantitative Analysis:** Success rates were calculated per problem and language group. (4) **Qualitative Analysis:** A post-survey was administered to capture user experience, analyzing themes regarding expressivity, trade-offs, and preference for English vs. Native language.

Study	Project	Setting	Design Goal	Participants	Methods
Rivero & Yin (2025)	Navigating the contradictions of AI in Bilingual Education	Tertiary-level teacher education program (USA)	Examine bilingual PSTs' perceptions of AI regarding their own biliteracy and pedagogical development.	13 Bilingual Pre-Service Teachers (PSTs) (8 Mandarin-English, 5 Spanish-English).	**Design-Based Intervention:** (1) **Reflective Journaling:** Participants completed journals reflecting on AI's affordances/constraints for their own literacy. (2) **Artifact Design:** PSTs designed and evaluated AI-generated multimodal texts for bilingual learners. (3) **AI Conversations:** Participants engaged in dialogic conversations with AI about bilingual education concepts. (4) **Analysis:** Reflexive thematic analysis was applied to journals and artifacts to identify tensions between AI support and standardization/monolingual ideologies.
Saddhono et al. (2024)	AI in Facilitating Multilingual Literacy and Cross-Cultural Understanding	Academic/ Educational Institutions	Evaluate the impact of AI tools (translation, learning apps, cultural platforms) on multilingual literacy and cultural competence.	500 students across various language backgrounds.	**Mixed-Method Study:** (1) **Group Assignment:** Students were assigned to groups using AI-Driven Language Learning (e.g., Duolingo), Automated Translation (e.g., Google Translate), or Cultural Education Platforms. (2) **Implementation:** Tools were used over one academic year. (3) **Data Collection:** Data included standardized test scores, reading/writing samples, and cultural understanding evaluations. (4) **Evaluation:** Pre-implementation and post-implementation scores were compared (e.g., literacy scores, engagement levels).
Zhang (2024)	Leveraging AI Tools for Discourse Analysis in Early Childhood	Early Childhood Bilingual Classroom (Elementary level, Chinese-English) in Arizona, USA	Explore how AI tools assist in discourse analysis and enhance bilingual skills in young children.	1 Bilingual Educator (managing a 5th-grade class of 30 students).	**Qualitative Case Study:** (1) **Contextual Training:** The educator received practical training on AI tools (ChatGPT, Gemini, Magic School, Canva) from the school district. (2) **Implementation:** The educator used these tools for lesson planning, creating gamified activities, and analyzing student discourse/responses. (3) **Data Collection:** A one-hour in-depth semi-structured interview was conducted with the educator. (4) **Analysis:** The interview was transcribed and thematically analyzed to identify benefits (e.g., personalized feedback) and challenges (e.g., privacy).

Tools discussion

The synthesized data demonstrates that AI tools, ranging from specialized platforms like **Promptly** and **Classroom Companion** to broad ecosystems like **ChatGPT, Gemini, and ELAI**, act as powerful catalysts for quantitative improvement in multilingual education. Across the studies, AI implementation consistently resulted in significant gains, most notably a **25–31% increase in literacy and vocabulary retention** (Aravind et al., 2025; Zhang, 2024) and a substantial rise in academic success rates (up to **77.8%** (Cole, 2025)). These tools successfully transition the learning experience from a "one-size-fits-all" model to a personalized one, where **real-time feedback** and **L1 (native language) prompting** lower the barrier to entry for complex subjects like computer science and academic writing (Khan et al., 2025; Rivero & Yin, 2025) Specifically, AI-integrated writing support within bilingual programs, has been shown to significantly enhance written expression skills by allowing students to bridge linguistic gaps during the drafting process (Krstinić & Bačić, 2025).

However, the chart also reveals a critical **socio-technical paradox**: while AI increases efficiency and student engagement (up to **46%** (Zhang, 2024)), it introduces new qualitative risks. These include "metacognitive laziness"(Khan et al., 2025), a loss of "emotional engagement" in virtual instructors (Liaqat & Munteanu, 2020), and a persistent **algorithmic bias** where low-resource languages like Arabic yield lower technical accuracy than English (Rivero & Yin, 2025). Furthermore, the "whitewashing" of cultural nuances (Saddhono et al., 2024) suggests that while AI is an excellent tool for **scaffolding and administrative relief** (saving teachers up to 41% of their time (Zhang, 2024)), it cannot yet replace the human-centered need for cultural authenticity and critical inquiry. Ultimately, the successful integration of AI in multilingual contexts depends on a balance between leveraging its **computational efficiency** and mitigating its **linguistic and cultural distortions**.

Table 2. Comparative Analysis of AI Implementations and Linguistic Outcomes

Study	AI Tool(s) Used	Goal of Using AI	Result of AI Implementation
Aravind et al. (2025)	**Adaptive AI System** (integrating NLP chatbots, Machine Learning algorithms, and Gamification)	To accelerate vocabulary and syntax mastery by providing personalized learning paths and real-time feedback adapted to the learner's pace.	• **Vocabulary Retention:** Improved by **25%** in the AI group vs. 12% in the control group. • **Syntax Accuracy:** Improved by **30%** in the AI group vs. 10% in the control group. • **Engagement:** Scores were significantly higher for the AI group due to interactive elements.
Cole (2025)	**Classroom Companion**	To provide immediate, detailed feedback on written assignments (grammar, syntax) and manage teacher grading workload in a high-pressure exam context.	• **Success Rate:** Increased from **54.8%** (Jan) to **77.8%** (Feb). • **Efficiency:** Average response time dropped from 9 to 8 minutes. • **Quality:** Student writing showed improved clarity and structure; teachers could target specific class-wide weaknesses.
Khan et al. (2025)	**ChatGPT**	To act as an on-demand writing assistant for brainstorming, structuring, and refining academic texts to reduce writing anxiety and improve fluency.	• **Fluency & Confidence:** Students reported enhanced writing flow and reduced "revision anxiety". • **Efficiency:** Improved time management in drafting tasks. • **Cognitive Risk:** Concerns arose regarding "metacognitive laziness" and over-reliance on the tool for critical thinking.
Miranda & Vegliante (2025)	**ELAI** (AI video generation) & **Google Translate**	To create multilingual e-learning courses with virtual speakers in the learners' native languages (Italian, Turkish, Polish, Portuguese) to improve accessibility.	• **Preference: 80%** of users preferred AI content in their native language over human speakers with subtitles. • **Effectiveness: 83%** found the AI-generated videos effective for learning. • **Knowledge:** Participants showed clear improvement between initial and final assessments.
Prather et al. (2025)	**Generative AI** (via **"Promptly"** tool)	To allow non-native English speakers to solve coding problems using prompts in their native languages (Portuguese, Chinese, Arabic).	• **Success Rates:** High success for Portuguese/ Chinese prompts; lower for Arabic due to model bias. • **Trade-off:** Students found native language prompting more "expressive" but admitted English prompts often yielded better technical accuracy.
Rivero & Yin (2025)	**ChatGPT (v4.0)**	To scaffold academic literacy and create multimodal texts (e.g., poems) while critiquing AI's "standardized" and "monolingual" biases.	• **Affordance:** AI successfully supported brainstorming and "polishing" grammar. • **Constraint:** AI frequently distorted cultural meanings (e.g., mistranslating "taste of home"), prompting students to develop a critical stance toward "whitewashed" outputs.
Saddhono et al. (2024)	**Duolingo, Babbel** (Learning); **Google Translate, DeepL** (Translation); **CultureGrams**	To facilitate multilingual literacy and cross-cultural understanding through personalized instruction and automated translation.	• **Literacy:** Multilingual Literacy Scores increased by **31%** (from 58 to 76). • **Engagement:** Student engagement levels rose by **46%**. • **Admin:** Teachers saved **41%** of their time on administrative tasks.

Study	AI Tool(s) Used	Goal of Using AI	Result of AI Implementation
Zhang (2024)	**ChatGPT, Gemini, Magic School, Canva**	To analyze student discourse for deeper insights into learning gaps, generate feedback, and create gamified/visual activities.	• **Insights:** AI provided "unexpected insights" into student learning that manual analysis missed. • **Expression:** Enhanced students' ability to express complex ideas in both Chinese and English. • **Engagement:** Gamified activities increased student motivation to participate.
Konca et al. (2025)	**Generative AI** (General adoption context)	To investigate factors influencing the *behavioral intention* of pre-service teachers to adopt AI in their future classrooms.	• **Adoption Drivers:** Validated that **Subjective Norms** (peer influence), **Self-Efficacy**, and **Anxiety** significantly predict the intention to use AI. • **Cultural Variance:** Anxiety negatively impacts "Perceived Usefulness" differently across cultures (Turkey vs. UAE).

DISCUSSION

Step 1: Understand and Specify the Context of Use

The context for Human-Centered Design in AI educational technologies is marked by a tension between traditional pedagogical structures and the specific needs of diverse, often marginalized, learner groups. **Aravind et al.**(Aravind et al., 2025) identify a context in Indian schools and language centers for children aged 5–12 where "one-size-fits-all" instruction leads to disengagement and poor retention of vocabulary and syntax due to passive rote memorization. In contrast, **Cole** (Cole, 2025) focuses on a post-COVID bilingual secondary school in Paris, where 16 to 17-year-old students (11th grade) face a "downward trend" in grammar mastery and high pressure for Cambridge/BFI exams, leading to a high temptation for AI-driven plagiarism. **Khan et al.** (Khan et al., 2025) explore the lived experiences of Indian EFL learners in higher education who face resource constraints and use ChatGPT as a cognitive scaffold to navigate the complexities of academic writing in a second language. **Konca et al.** (Konca et al., 2025) provide a macro-level context comparing Faculties of Education in Turkey and the UAE, noting that the UAE's collectivist, government-driven approach contrasts with Turkey's decentralized, individualistic landscape, where high power distance influences AI adoption. For mature immigrant English Language Learners (ELLs) in Canada, **Liaqat and Munteanu** (Liaqat & Munteanu, 2020) describe an informal learning environment where highly educated but underemployed users face "de-skilling," social isolation in "ethnic bubbles," and the stigma of illiteracy despite high tech-proficiency. **Miranda and Vegliante**(Miranda & Vegliante, 2025) identify the high-stakes European mining sector (Türkiye, Poland, Portugal, Italy), where miners and rescue members with low English proficiency require technical vocabulary in their native languages for safety, a need currently hampered by the high cost of human speakers. **Prather et al.** (Prather et al., 2024) investigate Non-Native English Speakers (NNES) in Computer Science across Portugal, New Zealand, and Saudi Arabia, where the "English-centric" nature of programming creates a massive cognitive load and feelings of embarrassment. **Rivero and Yin** (Rivero & Yin, 2025) focus on a USA graduate-level course with 13 bilingual pre-service teachers (PSTs), navigating a context where AI often reinforces "monolingual ideologies" and "standardized English," potentially erasing their linguistic identities. **Saddhono et al.** (Saddhono et al., 2024) examine 500 students in the "digital economy" facing language walls and a lack of access to high-quality cultural resources. Finally,

Zhang (Zhang, 2024) details the context of an Arizona Chinese-English immersion classroom where a single educator manages 30 students, struggling to assess the "depth" of student discourse and language acquisition progress in real-time.

Step 2: Specify the User Requirements

User requirements derived from these contexts focus on balancing technical efficiency with psychological safety and cultural relevance. In aligning with the contexts identified in Step 1, **Aravind et al.** (Aravind et al., 2025) determine that systems must provide personalized learning paths, real-time difficulty adaptation, and gamification to maintain engagement for young learners. **Cole** (Cole, 2025) specifies that students need bi-monthly writing practice and immediate feedback based on specific BFI rubrics, while teachers require a mechanism to prevent plagiarism and data to target class-wide weaknesses. **Khan et al.** (Khan et al., 2025) use Cognitive Load Theory (CLT) and the Technology Acceptance Model (TAM) to specify requirements for a "non-judgmental space," 24/7 tutoring support (AI-IDLE), and tools that reduce "extraneous cognitive load" like error identification. **Konca et al.** (Konca et al., 2025) utilize the GETAMEL framework, specifying that adoption depends on satisfying "Subjective Norms," "Perceived Usefulness," and addressing "Anxiety" related to self-efficacy. For mature immigrants, **Liaqat and Munteanu** (Liaqat & Munteanu, 2020) specify that formative feedback is more vital than numerical grades, requiring systems that foster community and provide "localized, specific guidance" to build reviewing skills. **Miranda and Vegliante**(Miranda & Vegliante, 2025) emphasize accessibility, requiring content delivered in native languages that is cost-effective, scalable, and visually appealing for asynchronous Moodle learning. **Prather et al.** (Prather et al., 2024) specify a need for "expressivity," allowing students to use their native logic (L1) to generate code without being hindered by English syntax memorization. **Rivero and Yin** (Rivero & Yin, 2025) require the development of a "critical disposition" toward AI, including pedagogical strategies like "socio-analytic artifacts" to critique algorithmic bias and linguistic erasure. **Saddhono et al.** (Saddhono et al., 2024) specify "real-time access" across languages and tools that foster "cultural sensitivity," while **Zhang** (Zhang, 2024) requires tools that provide "actionable insights" into student discourse and automate administrative tasks to allow for deeper analysis of language gaps.Beyond pedagogical feedback, a critical user requirement for high-stakes exam preparation is **modal consistency**; the digital environment must align with the physical constraints of the final assessment (e.g. handwriting) to prevent technical artifacts like typos from skewing the evaluation of linguistic proficiency(Cole, 2025).

Step 3: Produce Design Solutions to Meet These Requirements

Design solutions manifest as a mix of new software, pedagogical frameworks, and integrated toolkits. **Aravind et al.**(Aravind et al., 2025) developed an "Adaptive AI-Based Personalized Learning System" integrating NLP-based chatbots as virtual trainers and ML algorithms that adjust lesson difficulty based on performance. **Cole** (Cole, 2025) integrated the "Classroom Companion" AI program, which uses timed, prompt-based writing activities (e.g., analyzing *Intimate Apparel*) and allows students three revision attempts based on AI critique. **Khan et al.** (Khan et al., 2025) did not build a tool but designed a "usage workflow" where ChatGPT acts as an on-demand writing assistant for brainstorming and professional simulation. **Konca et al.** (Konca et al., 2025) designed a 25-item GETAMEL-based survey instrument, translated into Turkish and Arabic, to measure how cultural factors influence AI ease of use. **Liaqat and Munteanu**(Liaqat & Munteanu, 2020) designed a web-based "Peer-Review App" where users "tag" text (e.g., "express regret") to generate "localized rubrics" for peer feedback. **Miranda and Vegliante** (Miranda & Vegliante, 2025) built a content management system using Google Sheets and Google Script integrated with the **ELAI API** to generate AI virtual speakers in multiple languages (Italian, Turkish, Polish, Portuguese) for Moodle. **Prather et al.** (Prather et al., 2024) implemented "Promptly," a tool for "Prompt Problems" where students write natural language prompts (L1 or mixed) to generate code via LLMs and iteratively refine them based on test case outputs. **Rivero and Yin** (Rivero & Yin, 2025) designed "Design-Based Interventions" involving reflective journaling and the creation of multimodal artifacts (poems, images) to challenge AI definitions of bilingual education. **Saddhono et al.**(Saddhono et al., 2024) deployed an ecosystem including **Duolingo, Babbel, and Google Translate**, alongside "Cultural Education Platforms" to simulate immersion. **Zhang** (Zhang, 2024) utilized a toolkit of **ChatGPT** (thematic analysis), **Gemini** (reading comprehension feedback), **Magic School** (gamified activities), and **Canva** (visual aids) to structure sentence acquisition. The design solutions identified across these studies demonstrate a shift from purely functional tools to integrated pedagogical scaffolds. In the case of **Khan et al.**(Khan et al., 2025), the solution involves a usage workflow where AI models tone and format, bridging the gap between current proficiency and target performance through professional simulation. **Konca et al.**(Konca et al., 2025) expanded the definition of a design solution by developing a validated 25-item GETAMEL-based instrument, providing a blueprint for how AI training must be tailored to specific cultural norms like "Subjective Norms".

For informal learning, **Liaqat and Munteanu** (Liaqat & Munteanu, 2020) implemented a peer-review workflow that uses student-generated "tags" to create localized rubrics, moving away from generic feedback toward specific, scaffolded inquiry. Similarly, the ecosystem described by **Saddhono et al.**(Saddhono et al., 2024) combines adaptive applications with translation systems and immersion platforms to personalize content delivery via machine learning algorithms. Finally, **Zhang**(Zhang, 2024) illustrates a multi-tool approach where ChatGPT and Gemini are utilized for thematic analysis and feedback, allowing educators to generate "story starters" that structure complex sentence acquisition in bilingual contexts.

Step 4: Evaluate the Designs Against Requirements

Evaluations across these studies employ mixed-methods to measure both performance gains and social impacts. Aravind et al.(Aravind et al., 2025) used pre-tests and post-tests to show a 25% increase in vocabulary recall and a 30% increase in syntax accuracy, while qualitative analysis confirmed that gamified, adaptive elements successfully maintained learner interest compared to the control group. Cole(Cole, 2025) evaluated student data from January to February 2025, showing a success rate increase from 54.8% to 77.8%, though noting a mismatch between platform typing and handwritten exams. Khan et al. (Khan et al., 2025) used thematic analysis of interviews to confirm enhanced fluency, but warned of "metacognitive laziness" and loss of critical thinking. Konca et al.(Konca et al., 2025) tested their model via Structural Equation Modeling (SEM) and Confirmatory Factor Analysis (CFA) with 846 participants, finding that "Anxiety" negatively impacted "Perceived Usefulness" differently in the UAE versus Turkey (R^2=0.54 vs 0.49). Liaqat and Munteanu (Liaqat & Munteanu, 2020) conducted an "in-the-wild" deployment with 16 participants, finding that localized rubrics yielded 51.8% more relevant feedback than general ones, though engagement dropped without accountability. Miranda and Vegliante(Miranda & Vegliante, 2025) used UEQ and TUXEL models with 147 participants, showing 80% preferred native-language content but 17% missed the "emotional engagement" of human voices. Prather et al. (Prather et al., 2024) analyzed 1,679 prompts, revealing a bias where Portuguese and Chinese prompts were more successful than Arabic, and students often reverted to English for better technical accuracy. Rivero and Yin (Rivero & Yin, 2025) evaluated PST journals, finding that while AI helped academic literacy, it often "distorted" cultural meanings like "taste of home," yet the intervention successfully fostered critical evaluator roles. Saddhono et al. (Saddhono et al., 2024) showed a 31% increase in Multilingual Literacy Scores and a 41% reduction in teacher administrative time. Zhang (Zhang, 2024) used a qualitative case study to show that while AI provided "unexpected insights" into student thinking, it raised privacy risks and concerns over student over-reliance.

The evaluation of the 'Classroom Companion' revealed a significant socio-technical mismatch: while the AI successfully improved success rates to 77.8% and increased efficiency, the shift from digital typing to manual handwriting created a performance discrepancy. This suggests that HCD for exam-focused AI must account for the physicality of the end-use environment, as students were occasionally penalized for digital typos that do not reflect their true grammatical mastery in a handwritten context (Cole, 2025).

Furthermore, the evaluation by Rivero and Yin(Rivero & Yin, 2025) demonstrated that such interventions can move learners from passive users to critical evaluators capable of identifying when AI reproduces 'standard language ideologies.' In broader pedagogical settings, Saddhono et al. (Saddhono et al., 2024) confirmed that while literacy scores rose by 31% and engagement by 46%, the technology still struggles with 'cultural nuance' and idiomatic expressions. Finally, Zhang (Zhang, 2024) noted that while AI tools provide 'unexpected insights' into learning processes that manual review might miss, designers must remain vigilant regarding privacy risks and the long-term impact of student over-reliance on automated scaffolds.

Principal Findings

The systematic review of the selected literature reveals a paradigm shift in how Human-Centered Design (HCD) is applied to multilingual educational technologies. By analyzing the 10 core studies through the lens of ISO 9241-210, three primary themes emerge regarding the current state of the field:

1. From Content Delivery to Pedagogical Scaffolding

While early educational tools focused on static content delivery, current HCD practices prioritize **adaptive scaffolding**. Solutions such as the "Classroom companion" (Cole, 2025) and "Promptly" (Rivero & Yin, 2025) shift the AI's role from a text generator to a feedback mechanism. These designs meet the user requirement for "non-judgmental" spaces, allowing learners to iteratively refine their work. This transition successfully reduces the "extraneous cognitive load" (Khan et al., 2025) often associated with language barriers, moving the learner's focus from rote syntax memorization to higher-order problem-solving and logic.

2. The Tension Between Global Models and Local Nuance

A critical finding across the "Context of Use" and "Evaluation" phases is the persistent tension between the efficiency of Large Language Models (LLMs) and the preservation of cultural identity. While AI-driven tools demonstrated significant quantitative gains, such as a **31% increase in literacy scores** (Zhang, 2024) and **30% better syntax accuracy** (Aravind et al., 2025), qualitative evaluations revealed a "whitewashing" effect (Saddhono et al., 2024) In many cases, users experienced a "trade-off" where they reverted to English-centric prompting because the AI performed poorly in their native (L1) languages(Rivero & Yin, 2025), highlighting a systemic bias in current technological infrastructures.

3. Socio-Technical Barriers and Critical Agency

The evaluation of these HCD interventions suggests that technical success does not always equate to pedagogical health. Several studies noted risks of "metacognitive laziness" (Khan et al., 2025) and a decline in engagement due to lack of accountability in automated systems (Prather et al., 2024). Consequently, a new user requirement has emerged: **Critical AI Literacy**. Findings suggest that the most effective design solutions are those that move users from passive consumers to "critical evaluators" (Saddhono et al., 2024) who can identify algorithmic bias and linguistic erasure, ensuring that HCD supports, rather than replaces, culturally responsive teaching.

4. The Design-Reality Gap

Physicality and Modal Consistency A critical interaction design challenge identified is the lack of modal consistency between digital learning environments and physical assessment constraints . While AI platforms improve success rates and efficiency, a significant performance discrepancy emerges when students transition from digital typing to manual handwriting during high-stakes exams . This suggests that HCD for educational AI must account for the **physicality of the end-use environment**, ensuring that technical artifacts like digital typos do not unfairly penalize a student's true linguistic proficiency.

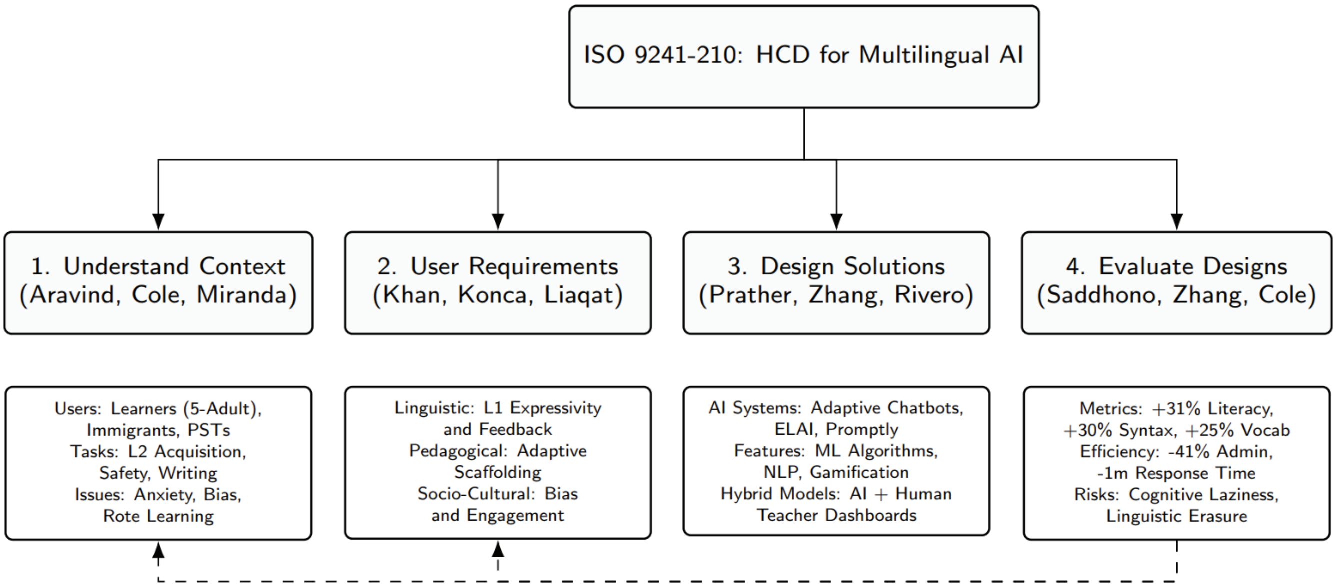

Figure 2. Synthesis of AI-Based Educational Technologies across the ISO 9241-210 HCD Lifecycle

Limitations

A primary technical limitation is the disparity in model performance across languages. While AI allows for native-language (L1) "expressivity," studies showed that prompts in low-resource languages (e.g., Arabic) yielded lower technical accuracy than English or high-resource languages like Chinese (Rivero & Yin, 2025). This creates a "trade-off" where students are forced to revert to English to ensure the AI functions correctly, paradoxically reinforcing the "English-centric" bias the tools were meant to circumvent (Rivero & Yin, 2025; Saddhono et al., 2024).

Erosion of Critical Thinking and Agency: The "ease of use" provided by AI tools introduces significant behavioral risks. Evaluators identified a trend toward "metacognitive laziness," where students rely on AI for text generation rather than using it as a scaffold for their own logic (Khan et al., 2025). This is compounded by an over-reliance on automated feedback, which can lead to a decline in student engagement over time if there is no human accountability or instructor oversight to validate the AI's output (Konca et al., 2025; Prather et al., 2024).

Cultural and Emotional Distortion: AI tools frequently struggle with "cultural nuance" and idiomatic expressions, often mistranslating deep cultural concepts (e.g., "taste of home") or reinforcing "monolingual ideologies" that erase the diverse linguistic identities of bilingual students (Saddhono et al., 2024; Zhang, 2024). Furthermore, despite the efficiency of AI-generated voices and avatars, users reported a distinct lack of "emotional engagement" compared to human instructors, suggesting that current AI lacks the affective capabilities necessary for deep, empathetic language learning(Liaqat & Munteanu, 2020).

Infrastructure and Assessment Mismatches: There is a notable "design-reality gap" regarding how these tools are evaluated. For example, while AI platforms improve typing-based literacy and efficiency, they do not account for the physical requirements of high-stakes exams (e.g., handwriting), potentially penalizing students for typos or lack of manual writing speed(Cole, 2025). Additionally, the high demand for "24/7 tutoring" and real-time data analysis places a hidden strain on privacy and data security that current educational frameworks are often unequipped to handle. (Konca et al., 2025)

Socio-Technical Barriers: The effectiveness of AI is heavily mediated by cultural and psychological factors that vary by region. Requirements like "Subjective Norms" and "Anxiety" significantly impact how teachers and students adopt these tools; a design that works in a decentralized system (Turkey) may fail in a collectivist, government-driven one (UAE) because the underlying social motivations for using technology differ. (Miranda & Vegliante, 2025)

CONCLUSION

This systematic review demonstrates that while AI-driven tools, ranging from adaptive chatbots and LLM-based coding assistants to multilingual virtual avatars, significantly enhance linguistic fluency, vocabulary recall (up to 31%), and classroom efficiency, their success is strictly contingent upon a **Human-Centered approach**. Effective design must move beyond "one-size-fits-all" instruction to address specific cultural anxieties , cognitive loads , and the "de-skilling" of immigrant learners.

A critical tension persists in the current landscape: while AI facilitates "expressivity" by allowing students to use their native languages (L1) as a logical scaffold the technology frequently reinforces "monolingual ideologies" and struggles with deep cultural nuances. This necessitates a **design shift** where AI is utilized not merely for static content generation, but as a dynamic tool for iterative feedback and critical inquiry . The evolution of HCD in this field indicates that AI is moving from a simple "translator" to a "pedagogical scaffold" , requiring a focus on feedback loops rather than just content delivery.

Ultimately, the integration of AI in multilingual settings effectively bridges resource gaps and reduces administrative burdens, saving teachers up to 41% of their time. However, to address the **"Linguistic Power Gap"** and the risk of "linguistic erasure" favoring English, a **"Critical-Dynamic Paradigm"** is required. This paradigm calls for tools built to protect local cultural nuances rather than enforcing standardized English molds.

Furthermore, the **"Human-in-the-Loop" necessity** remains paramount; AI cannot function in a vacuum. Most successful HCD interventions are **hybrid models** that leverage AI for administrative efficiency while relying on human educators for emotional engagement and critical thinking. This is particularly evident in contexts where engagement decay occurs; while AI-driven localized rubrics initially generate 51.8% more relevant feedback, user motivation often wanes without the institutional accountability and validation provided by a human instructor.

Declaration

Conflict of interest: The authors declared no potential conflicts of interest for the research, authorship, or publishing of this work.

Ethical statements: We hereby declare that this manuscript is the result of my independent creation under the reviewers' comments. Except for quoted contents, this manuscript does not contain any research achievements that have been published by other individuals or groups.

REFERENCES

Aravind, Angalakuduru, M. Durairaj, Preeti Chitkara, et al. "Adaptive AI-Based Personalized Learning for Accelerated Vocabulary and Syntax Mastery in Young English Learners." *International Journal of Advanced Computer Science and Applications* 16, no. 4 (2025).

Botarleanu, Robert-Mihai, Micah Watanabe, Mihai Dascalu, Scott A. Crossley, and Danielle S. McNamara. "Multilingual Age of Exposure 2.0." *International Journal of Artificial Intelligence in Education* 34, no. 4 (2024): 1353–77. https://doi.org/10.1007/s40593-023-00386-7.

Chen, Xiaojiao, Zhebing Hu, Yuanyuan Li, and Chengliang Wang. "The Journey of Challenges and Triumphs: A Systematic Literature Review of the Dynamic Evolution of Human-Centered Artificial Intelligence in Education." *Interactive Learning Environments* 33, no. 8 (2025): 4906–31. https://doi.org/10.1080/10494820.2025.2472288.

Cole, Jennefer York. "Using AI in the Bilingual Classroom to Enhance Writing Skills in Second Language Learning: A Case Study." *Journal of Education, Innovation and Communication* 7, no. 3 (2025): 51–58. https://doi.org/10.34097/jeicom-7-3-5.

Dr. Nishat Zafar, Dr. Saira, and Seerat Afzal. "AI-Powered Reading Support for Multilingual Learners in Higher Education: A Critical Review." *Journal for Social Science Archives* 3, no. 1 (2025): 776–86. https://doi.org/10.59075/jssa.v3i1.158.

Du, Mingxuan, Lili Zhu, and Mario Di Nardo. "Strategic Application of Deep Learning Methods in Global Educational Collaboration." *Proceedings of the 2025 International Conference on Big Data and Informatization Education*, February 21, 2025, 313–17. https://doi.org/10.1145/3729605.3729660.

Hamid, Shazia. *Integrating Artificial Intelligence and Multimodality in Language Education: A Systematic Review of Emerging Trends and Practices*. 4, no. 2 (2025).

ISO. "ISO 9241-210:2019." Accessed March 16, 2026. https://www.iso.org/standard/77520.html.

Khan, Razia, Md. Tauseef Qamar, Mohd Sajid Ansari, and Juhi Yasmeen. "Enhancing or Impairing? Exploring Indian EFL Learners' Academic Writing Narratives with ChatGPT." *Cogent Education* 12, no. 1 (2025): 2514329. https://doi.org/10.1080/2331186X.2025.2514329.

Konca, Ahmet Sami, Ahmet Simsar, Reem Alhajji, and Afra Al Mansoori. "Cross-Cultural Perspectives on AI Adoption in Teacher Education: A Comparative Study of Pre-Service Teachers in Turkey and the United Arab Emirates." *Interactive Learning Environments* 33, no. 10 (2025): 5820–46. https://doi.org/10.1080/10494820.2025.2488143.

Krstinić, Marija, and Lucija Bačić. *Understanding the Language Gap in Artificial Intelligence*. 19.12.2025.

Liaqat, Amna, and Cosmin Munteanu. "Leveraging Peer Support for Mature Immigrants Learning to Write in Informal Contexts." *Proceedings of the ACM on Human-Computer Interaction* 4, no. CSCW2 (2020): 1–24. https://doi.org/10.1145/3415177.

Miranda, Sergio, and Rosa Vegliante. "Leveraging AI-Generated Virtual Speakers to Enhance Multilingual E-Learning Experiences." *Information* 16, no. 2 (2025): 132. https://doi.org/10.3390/info16020132.

Mirnig, Alexander G., Alexander Meschtscherjakov, Daniela Wurhofer, Thomas Meneweger, and Manfred Tscheligi. "A Formal Analysis of the ISO 9241-210 Definition of User Experience." *Proceedings of the 33rd Annual*

ACM Conference Extended Abstracts on Human Factors in Computing Systems, April 18, 2015, 437–50. https://doi.org/10.1145/2702613.2732511.

Özyurt, Sinan. "AI-Assisted English Language Learning for Cross-Cultural Medical Education in Multilingual Settings." *Experimental and Applied Medical Science* 5, no. 2 (2024): 83–92. https://doi.org/10.46871/eams.1464830.

Prather, James, Brent N. Reeves, Paul Denny, et al. "Breaking the Programming Language Barrier: Multilingual Prompting to Empower Non-Native English Learners." arXiv:2412.12800. Preprint, arXiv, December 17, 2024. https://doi.org/10.48550/arXiv.2412.12800.

Rivero, Edward, and Peng Yin. "Navigating the Contradictions of AI: Critiquing AI's Standardized English and Developing Creative Bilingual Possibilities." *Frontiers in Education* 10 (July 2025): 1616458. https://doi.org/10.3389/feduc.2025.1616458.

Saddhono, Kundharu, Muhammad Rohmadi, Try Annisa Lestari, Kaleb E. Simanungkalit, and Indriyo K. Sukmono. "The Role of AI in Facilitating Multilingual Literacy and Cross-Cultural Understanding." *2024 International Conference on IoT, Communication and Automation Technology (ICICAT)*, November 23, 2024, 366–71. https://doi.org/10.1109/ICICAT62666.2024.10923254.

Saini, Mohit. *Educators and AI in Collaboration: Enhancing Multilingual Teaching and Learning in Higher Education Through Natural Language Processing*. 3, no. 1 (2025).

Schmager, Stefan, Ilias O. Pappas, and Polyxeni Vassilakopoulou. "Understanding Human-Centred AI: A Review of Its Defining Elements and a Research Agenda." *Behaviour & Information Technology* 44, no. 15 (2025): 3771–810. https://doi.org/10.1080/0144929X.2024.2448719.

Shoeibi, Nastaran, Roberto Therón Sánchez, and Francisco José Garcia-Peñalvo. *Evaluating the Effectiveness of Human-Centered AI Systems in Education*. 2023.

Virvou, Maria. "Artificial Intelligence and User Experience in Reciprocity: Contributions and State of the Art." *Intelligent Decision Technologies* 17, no. 1 (2023): 73–125. https://doi.org/10.3233/IDT-230092.

Youness Moudettir, Siham Ouhrir, and Said Lotfi. "Enhancing Cognitive Ergonomics and Pedagogical Quality a Comprehensive Evaluation Framework for Technological Pedagogical Tools in Physical Education." *Cuestiones de Fisioterapia* 54, no. 3 (2025): 5007–29. https://doi.org/10.48047/53m9x307.

Zhang, Xinyu. "Leveraging AI Tools for Discourse Analysis in Early Childhood Bilingual Education: Enhancing Language Development." *Journal of Education and Educational Policy Studies* 2, no. 1 (2024): 24–32. https://doi.org/10.54254/3049-7248/2/2024011.